HEURISTICS IN CLINICAL JUDGMENT AND THE LABELING THEORY OF MENTAL ILLNESS

By

CURTIS CRAIG WALLING

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1987

To my parents.

Only their support has made my doctorate possible.

ACKNOWLEDGMENTS

I would like to thank the members of my doctoral committee. As chair of my committee, Greg Neimeyer contributed heavily toward my professional development as a counseling psychologist. I appreciate Jim Archer, Mary Fukuyama, Paul Fitzgerald, and Mark Alicke for graciously lending so much of their time and expertise as committee members.

I would also like to thank John Twomey for believing in me and helping me to believe in myself. Ces Bibby deserves mention for her good-natured assistance with many necessary details.

Finally, I must thank my late friend Jeff Ackerman and my late grandmother Lillian Cunningham for their encouragement and personal support. Both have been sorely missed by myself and others.

TABLE OF CONTENTS

LIST OF TABLES

Abstract of Dissertation Presented to the Graduate School of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

HEURISTICS IN
CLINICAL JUDGMENT AND THE LABELING
THEORY OF MENTAL ILLNESS

By

Curtis Craig Walling

May 1987

Chairman: Dr. Greg Neimeyer
Major Department: Psychology

According to labeling theory, mental illness does not exist intrapsychically but is created when clinicians label a person "mentally ill." Studies identifying labels assigned to patients and the clinical experience of judges as sources of bias in clinical judgment have been interpreted as support for labeling theory. Generalization from the literature on socio-cognitive biases, however, suggests that clinicians' utilization of judgmental heuristics may have contributed to these effects. The present study hypothesized that the anchoring heuristic contributes to labeling bias and clinical experience effects. Support for this hypothesis would qualify labeling theory and suggest instead that clinicians are subject to the same cognitive errors and biases as laymen.

American Psychological Association Division 29 members (N = 193) sampled by mail and undergraduate abnormal psychology students (N = 80) read a brief case study and evaluated the client's level of pathology and prognosis. The client was assigned one of two diagnostic labels: Adjustment disorder with depressed mood or Unipolar depression. Pretesting demonstrated that these labels yielded different ratings of pathology and prognosis but were equally representative of the client. The time of presentation of the labels was varied as the anchoring factor.

Results of a series of 2 (label) x 2 (experience) x 2 (time) analysis-of-variance procedures revealed a significant effect for labeling on level of pathology ratings. This finding replicates previous studies and extends them to include diagnoses as potential sources of bias. However, no other effects or interactions were significant, and the contribution of anchoring to labeling bias and clinical experience effects was undetermined. It was concluded that failure to control the representativeness of labels to stimulus subjects in previous studies may have contributed to the labeling bias and clinical experience effects found.

A method of obtaining significant anchoring effects, further research investigating the role of heuristics in labeling and other theories, and the necessity of demonstrating the representativeness of labels with stimulus subjects in analogue research were discussed.

CHAPTER I

INTRODUCTION

Clinical judgment has been defined as "an important human cognitive activity, typically carried out by a professional person, aimed at the prediction of significant outcomes in the life of another individual" (Goldberg, 1968, p. 483). Clinicians are persons who make an important contribution to our society. Unfortunately, reviewers have concluded that clinicians' judgments can be notoriously inaccurate (Arkes, 1981; Dawes, 1976; Einhorn & Hogarth, 1978; Kruglanski & Ajzen, 1983; Lourens, 1979; Meehl, 1954; Sawyer, 1966). Furthermore, extreme confidence often accompanies erroneous judgments (Fischoff, Slovic, & Lichtenstein, 1977). A major goal of clinical judgment research has been to identify factors that impede clinicians' accuracy (Heverly, Fitt & Newman, 1984). The present study is an attempt to extend investigations of some of the sources of inaccuracy in clinical judgment.

A clinician asked to evaluate a patient is faced with far more information than he or she can possibly process (Bieri, Atkins, Briar, Leaman, Miller, & Tripodi, 1966; Slovic & Lichtenstein, 1971). Assigning a label to a patient allows a clinician to categorize and manage this

overload of information (Langer & Abelson, 1974; Neisser, 1967). As a result, labeling has traditionally been recognized as an indispensable first step toward responsible treatment and routinely practiced throughout the mental health field (Braginsky, Braginsky, & Ring, 1969; Pepinsky & Pepinsky, 1954).

Although labeling serves an adaptive function by organizing input, researchers have long established that the use of a label creates a "set" that influences subsequent perceptions of the person or target labeled (Asch, 1946; Katz & Braly, 1947; Kelley, 1950; Mensh & Wishner, 1947). The methodology these researchers employed consisted of assigning various labels (e.g., "warm" and "cold") to a person or target and assessing the reactions of subjects to the labeling manipulation. Precautions were taken to ensure that the persons or targets differed only in the labels they were assigned. Subjects in these studies were found to consistently underevaluate or overevaluate the same person or target when differentially labeled. Thus, by assigning varying labels, researchers demonstrated as early as the 1940s that subjects can be led to perceive the same person or target in consistently different ways.

Extensions of these early studies have shown that clinicians are also susceptible to labeling bias effects. Temerlin (1968) varied the labels "neurotic" and "psychotic" and reported biased judgments of an audiotaped interview subject among psychiatrists, clinical psychologists, and

graduate students in clinical psychology, with psychiatrists more susceptible to labeling bias than the other two groups. Langer and Abelson (1974) varied the labels "job applicant" and "patient" and reported biased judgments of a videotaped interview subject by psychodynamically-oriented clinicians but not by behaviorally-oriented clinicians. Caetano (1974) introduced videotaped subjects as either individuals "off the street" or "patients in a state mental hospital" and reported more biased judgments of the subjects by psychiatrists than by students in an abnormal psychology class. Berman and Berman (1984) varied the labels "normal" and "psychotic" and reported biased judgments of a case study subject among social workers and graduate students in social work, with those individuals with higher levels of training more susceptible to labeling bias. These four studies suggest that labels and clinical experience within the medical model may be important sources of bias in clinical judgment.

These studies and others identifying labeling bias and clinical experience effects have been interpreted as evidence for a labeling theory of mental illness (Becker, 1963; Scheff, 1964, 1966, 1974). Labeling theory, also known as societal reaction theory, has generated considerable controversy since the 1960s by challenging the traditional psychiatric definition of mental illness. Briefly, labeling theory asserts that mental illness is not a quality of the individual, but rather a creation of social definitions

(Henry & Cohen, 1983). According to labeling theorists, clinicians' training encourages them to recognize "signs" of mental illness. The label "mental illness" is assigned by clinicians not to disease processes but to violations of social norms for which society provides no explicit label (Horwitz, 1979). Little significance is attached by labeling theorists to these violations. Rather, the act of being labeled mentally ill is viewed as the crucial step in the development of "mental illness." Labeling is seen as creating a self-fulfilling prophecy (Jones, 1977) where individuals who are labeled are treated as mentally ill and come to incorporate this image into their self-concepts (Jones, Farina, Hastorf, Markus, Miller, & Scott, 1984). The labeled individual is thus viewed as deeply stigmatized, with the identification of him or her as mentally ill leading to reactions such as rejection or isolation (Larsson & Starrin, 1983).

Numerous studies (Calicchia, 1981; Crumpton, Weinstein, Acker, & Annis, 1967; Franchia, Canale, Cambria, Ruest, & Sheppard, 1976) have confirmed that mental patients and ex-patients are generally perceived as threatening and socially undesirable. A review of studies on the public's attitude toward mental disorders concurs that "people who have experienced psychiatric problems are very much feared, disliked, and rejected" (Farina, 1981, p. 223). Stimulated by this stigma, labeling theorists have shifted attention away from labeled individuals to examine those persons

making judgments (clinicians), the process by which judgments are made, and the situations in which they are made (Henry & Cohen, 1983).

Although labeling theory has made important contributions (e.g., identifying the stigmatizing effects of diagnostic labels), its basic tenet, that clinicians' experience within the medical model and society's reaction to diagnostic labels are the primary etiological factors in the development of "mental illness," has been rejected by most researchers (Gove, 1982; Scull, 1984). Generalization from the literature on social and cognitive biases (Arkes, 1981; Friedlander & Stockman, 1983; Hamilton, 1981; Nisbett & Ross, 1980; Strasburger & Jackson, 1977; Strohmer, Haase, Biggs, & Keller, 1982; Tversky & Kahneman, 1974) suggests interpreting labeling bias and clinical experience effects as erroneous cognitive processes on the part of clinicians rather than manifestations of the propagation of "mental illness" where none actually exists. Therefore, the purpose of the present study was to provide support for the conceptualization of judgmental heuristics (Tversky & Kahneman, 1974) as cognitive processes that can contribute to erroneous clinical judgments. Specifically, the present study investigated whether the judgmental heuristic "anchoring" (Tversky & Kahneman, 1974) contributes to labeling bias and clinical experience effects in clinical judgment.

According to Tversky and Kahneman (Kahneman & Tversky, 1972, 1973; Tversky & Kahneman, 1971, 1973, 1974), when

people are faced with a judgment task they rely on anchoring and other simple heuristics (representativeness, availability) rather than sophisticated information-processing strategies. While these heuristics reduce the complexity of judgmental tasks and generally lead to reasonable judgments, research has indicated that utilizing them can lead to systematic errors and biases (Ajzen, 1977; Kahneman & Tversky, 1973; Nisbett & Borgida, 1975; Nisbett & Ross, 1980). While explanations of inaccurate clinical judgments are incomplete (Cline, 1985), Tversky and Kahneman's conceptualization of anchoring and other judgmental heuristics holds considerable promise (Ross, 1977; Wiggins, 1981).

Tversky and Kahneman suggest that the anchoring heuristic incurs bias when people make initial judgments and insufficiently adjust these judgments as additional information is processed, yielding an inaccurate final judgment. The stimulus which provides the initial judgment or starting point is called the "anchor." Using Tversky and Kahneman's theory of judgmental heuristics, Friedlander and Stockman (1983) demonstrated that anchoring effects occur in clinical judgment when they found that clinicians' estimates of pathology and prognosis were related differentially to the time of presentation (early vs. late) of pathological case material. In other words, pathognomonic information presented at the beginning of a case study served as an anchor to bias information processing and subsequent judgments.

Additionally, in a follow-up study to Friedlander and Stockman, Friedlander and Phillips (1984) found that clinicians may be more susceptible to these anchoring effects than students, and that clinicians were more confident of their judgments than were students.

The present study suggests that the Friedlander and Stockman (1983) and Friedlander and Phillips' (1984) findings parallel the previous studies (Berman & Berman, 1984; Caetano, 1974; Langer & Abelson, 1974; Temerlin, 1968) identifying labeling bias and clinical experience effects in clinical judgment. Labels presented at the outset of judgment tasks may have served as anchors in these previous studies, and experienced clinicians may have been more susceptible to anchoring effects than less experienced or inexperienced judges. It seems likely that a replication of these studies would be qualified by an anchoring manipulation. According to Levin, "anchoring may provide an explanation for the results of earlier studies that noted the rapid stabilization of clinical impressions (Meehl, 1960) and the 'rigidifying effect' (Dailey, 1952) of initial references" (Levin, 1984, p. 9). The present study concurs with this reasoning in applying an anchoring explanation, rather than labeling theory, to previous studies identifying clinical experience and labeling bias effects in clinical judgment.

It is unsurprising that previous studies manipulating labels such as "psychotic" and "normal" elicited different reactions in subjects. These studies utilized vastly disparate labels or even semantic opposites that were not demonstrated to be equally representative of their respective stimulus subjects. Since the representativeness, or goodness-of-fit, of these labels was not controlled, differential representativeness may have contributed to the labeling bias and clinical experience effects found. The present study suggests that internal validity would be enhanced by utilizing labels previously demonstrated to equally represent the case study chosen for investigation.

The present study operationally defines "clinical judgment" as clinicians' estimates of a case study subject's level of pathology on the Global Assessment Scale (GAS; Endicott, Spitzer, Fleiss, & Cohen, 1976), and prognosis on the Prognostic Scale (Friedlander & Stockman, 1983). These two dependent measures were selected as operational definitions of clinical judgment by Friedlander and Stockman (1983), Friedlander and Phillips (1984), and Levin (1984), based on Pepinsky and Pepinsky's (1954) assertion that judgments of pathology and prognosis are critical elements of effective treatment planning. Six hypotheses were tested:

1. Assigning varying diagnostic labels (Adjustment disorder with depressed mood or Unipolar depression) to a case study subject will affect evaluations of the subject's

level of pathology on the GAS and prognosis on the Prognostic Scale.

2. The level of clinical experience of the evaluators (clinician or student) will affect their evaluations of the case study subject's level of pathology on the GAS and prognosis on the Prognostic Scale.

3. Varying the time of presentation of the diagnostic labels (early or late) in the case study will affect evaluations of the subject's level of pathology on the GAS and prognosis on the Prognostic Scale.

4. The diagnostic labels assigned to the case study subject will interact with the time of presentation of the labels to affect evaluations of the subject's level of pathology on the GAS and prognosis on the Prognostic Scale.

5. The level of clinical experience of the evaluators will interact with the time of presentation of the diagnostic labels in the case study to affect evaluations of the subject's level of pathology on the GAS and prognosis on the Prognostic Scale.

6. Clinicians will place more confidence than students in their ratings of level of pathology on the GAS and prognosis on the Prognostic Scale.

The present study is divided into five chapters. Chapter I has introduced the study and presented the hypotheses that were tested. Chapter II presents a review of the relevant literature. Chapter III outlines the research

methodology employed. Chapter IV reports the results of the study. Chapter V is a discussion of the results, including implications and limitations of the study and suggestions for continued research.

CHAPTER II

REVIEW OF THE LITERATURE

Asch (1946), in a classic study of impression formation, gave subjects a list of traits describing an unknown person and asked them to rate the person on a variety of measures. One group of subjects received the following list: intelligent, skillful, industrious, warm, determined, practical, and cautious. A second group received another list: intelligent, skillful, industrious, cold, determined, practical, and cautious. These two lists differed in only one word--warm versus cold. However, very different overall impressions were formed, i.e., subjects judged the person described by the first list much more generous and wise than the person described by the second list.

Kelley (1950) replicated Asch's study by labeling a lecturer as either warm or cold and found perceptions of the lecturer to be significantly affected in the predicted positive or negative direction. Classroom observers were more likely to interact with the lecturer when he was labeled warm. Huguenard, Sager, and Ferguson (1970) replicated Asch's warm-cold manipulation in a simulated employment interview and also varied the length of the interview. They found that postinterview ratings were predictably

affected by the interviewer's initial set (warm-cold), but not by the length of the interview. These studies reflect the strength and persistence of labels--the initial set they create is not easily overridden.

An inferential bias can be defined as an error stemming from an individual's perceptual and decision-making apparatus (Perloff, Padgett, & Brock, 1980). Following this early work by Asch and others, researchers have manipulated various labels and identified inferential biases that can result from their use (Arkes, 1981). The first section of this review of the literature summarizes the research investigating labeling bias that occurs in clinical judgment.

Labeling Bias

A study by Rosenhan (1973) is often cited as a provocative example of the biased treatment accorded those labeled with psychiatric diagnoses. Rosenhan was interested in investigating the efficacy of psychiatric diagnosis and hospitalization. Can sane individuals be distinguished from the insane? Do the characteristics of mental hospitals cause people to display abnormal behaviors? To attempt to answer these questions Rosenhan devised the following scheme.

Rosenhan and several confederates voluntarily sought admission to various psychiatric hospitals under assumed names, feigning that they were hearing voices. They did not

report any other abnormal behaviors. The object of the study was to see if normal persons would be admitted under these circumstances and how long it would take them to be released if admitted. All of these pseudopatients were diagnosed as schizophrenic save one, who was diagnosed manic-depressive, and all were admitted for hospitalization.

Following admission they stopped feigning that they were hearing voices and behaved as sanely as possible. Although their symptoms disappeared the diagnoses assigned to these pseudopatients remained influential, as none was immediately released. The average time until release was 19 days, with a range of 7 to 52 days. When these pseudopatients were finally released, all were assigned the diagnosis "schizophrenia in remission." None of the staff members at any of the psychiatric hospitals had detected their sanity. Rosenhan concluded "it is clear that we cannot distinguish the sane from the insane in psychiatric hospitals" (Rosenhan, 1973, p. 257).

Rosenhan's innovative scheme has inspired both acclaim and controversy. Cozby (1977) claims that Rosenhan's study is valuable because the data were collected in a real setting. According to Cozby, these data provide a meaningful interpretation of life in a mental institution and a provocative example of the stigmatizing effects of psychiatric diagnoses. However, Rosenhan's study has been repeatedly challenged on conceptual, methodological, and ethical grounds. Many researchers, most notably Spitzer

(1976), assert that Rosenhan's conclusion was too broad, the nonexperimental methodology he employed has questionable external validity, and observer bias may have been present. Additionally, serious ethical issues such as the deception of the hospital staff and the placing of experimental participants at risk have been raised, making replication unlikely. Thus, Rosenhan's indictment of psychiatric diagnosis is likely to remain startling but controversial.

Thomas Szasz and R.D. Laing are psychiatrists who have long been outspoken critics of psychiatric diagnosis. Szasz makes a case for dispensing with the term "mental illness." He argues that mental illness does not exist, except as a metaphor, and insists that true illness or disease can affect only the body. He suggests that psychiatric diagnoses were conveniently constructed by clinicians to resemble medical diagnoses and applied to persons whose behavior annoys or offends others. In his book The Myth of Mental Illness, Szasz (1961) claims that psychiatric diagnoses are not classifications of disease but stigmatizing labels applied to inappropriate social behavior.

Laing (1967) concurs with Szasz. Laing (1967) concludes that the behavior of individuals labeled mentally ill is not irrational but sensible given the devastating demands life has placed upon them. Laing and Esterson add in their book Sanity, Madness, and the Family (1964) that assigning a diagnostic label begins a self-fulfilling prophecy, where an individual internalizes the diagnostic stereotype and

behaves accordingly. Laing has authored several books in which he argues that labels of mental illness are invalid and debilitating and should not be used.

Rosenhan's classic field observation study and the popular theoretical contributions of Szasz and Laing suggest that utilizing psychiatric labels can stigmatize individuals and bias clinicians' judgments of them. The remainder of this section of the review of the literature summarizes the empirical research investigating labeling bias in clinical judgment.

According to Wechsler, Solomon, and Kramer (1970), the landmark early work identifying labeling bias was Nunnally's (1961) book Popular Conceptions of Mental Health. Nunnally summarized the results of a longitudinal research program assessing attitudes toward the mentally ill as follows:

> The most important finding from our studies of public attitudes is that the stigma is very general, both across social groups and across attitude indicators. There is a strong "negative halo" associated with the mentally ill. They are considered, unselectively, as being all things "bad." Some of the "bad" attitudes that people have toward the mentally ill are partially supported by the facts--for example, the mentally ill sometimes are unpredictable and dangerous. However, the average man generalizes to the point of considering the mentally ill as dirty, unintelligent, insincere, and worthless. (Nunnally, 1961, p. 233)

A review by Rothaus, Hanson, Cleveland, and Johnson (1963) concurred with Nunnally. Rothaus et al. noted that nine studies conducted between 1955 and 1960 demonstrated that a stereotype of mental illness existed that carried

very negative connotations. Rothaus et al. hypothesized that the negative stereotype associated with the mental illness label could be circumvented by describing patients in problem-centered language rather than language reflecting mental illness. They asked employment interviewers to conduct an employment interview with a patient. Prior to the interview, half the interviewers were given information concerning the patient's background suggesting interpersonal problems, the other half information suggesting mental illness. Not surprisingly, those interviewers set to view the patient in terms of interpersonal problems gave more positive postinterview ratings than those set to view the patient in terms of mental illness.

Temerlin (1968) had psychiatrists, clinical psychologists, and graduate students in clinical psychology evaluate an audiotape of an interview with a normal, healthy man. When the interviewee was introduced by a prestigious colleague as "a very interesting man because he looks neurotic, but actually is quite psychotic," 28% of the subjects diagnosed psychosis and 63% diagnosed neurosis or character disorder. Among subject groups, the psychiatrists were found to diagnose psychosis significantly more often than the clinical psychologists or graduate students. In control conditions where no introduction preceded the audiotape or where the introduction was reversed so that mental health was suggested, none of the subjects diagnosed psychosis and only 22% diagnosed neurosis or character

disorder. Temerlin called these results "suggestion effects in psychiatric diagnoses" (Temerlin, 1968, p. 349) and concluded that a greater awareness of such effects would enhance the accuracy and helpfulness of clinical judgment.

In an extension of his 1968 study, Temerlin (1970) added nonprofessional subject groups of law students and undergraduates. Clinicians were found to be more susceptible to suggestion effects than the law students or undergraduates, and among clinician groups psychiatrists were again found to be more susceptible to suggestion than the clinical psychologists or graduate students in clinical psychology. Temerlin (1968, 1970) identified three factors he believed may have contributed to the suggestibility of psychiatrists in these two studies. First, psychiatrists, due to their training as physicians, may tend to diagnose illness in uncertain diagnostic situations. Second, psychiatrists' status and role hierarchies may reward conformity with prestige figures. Third, psychiatrists probably encounter more psychotics than clinical psychologists and graduate students encounter, and thus may have expected a psychotic.

Temerlin concluded that his results did not mean that clinical training reduces diagnostic accuracy, but that clinicians may have a set to perceive mental illness and that the effect of this set may be increased by suggestibility. He further concluded "the susceptibility of psychiatric diagnosis to distortion through prestige suggestion,

in combination with the stigmatizing connotation of psychiatric diagnosis, suggests that traditional diagnostic labels should not be used" (Temerlin, 1970, p. 117).

Langer and Abelson (1974) had clinicians representing two schools of thought, behavioral and analytic, evaluate a videotaped interview of a man who had recently applied for a job. Before viewing the tape, the interviewee was described to the clinicians as either a "job applicant" or a "patient." The interviewee was rated as fairly well adjusted by the behavioral therapists, regardless of the label assigned. However, the analytic therapists rated the interviewee as significantly more disturbed when he was labeled a patient than when he was labeled a job applicant.

Langer and Abelson concluded that the labeling bias identified in their study and previously in the Rothaus et al. and Temerlin studies has unfortunate practical consequences: Once an individual is labeled a "patient" a clinician may attribute subsequent behaviors to the patient's "illness" rather than actual external factors. However, Langer and Abelson noted that their study did not identify why behavior therapists were apparently immune to the labeling manipulation, or why analytic therapists succumbed to it. They strongly suspected that circumstances could be arranged where behavior therapists would be equally susceptible to bias, emphasizing "no single type of orientation toward clinical training is likely to avoid all types of biases or blind spots" (Langer & Abelson, 1974, p. 9).

Caetano (1974) had psychiatrists and undergraduate students in an abnormal psychology class evaluate two videotaped psychiatric interviews, one of a mental patient and one of a student. Instructions preceding the two videotapes identified the interviewees as either individuals "off the street" who were "paid to participate" or as "patients in a state mental hospital." After viewing the videotapes subjects were asked to diagnose and describe the interviewees. When the interviewees were identified as mental patients they were generally rated more mentally ill than when they were identified as paid participants off the street. This labeling bias was more prominent among the psychiatrists than the abnormal psychology students. Caetano concluded that clinical experience and the suggestion of mental illness provided by labels can bias psychiatric diagnosis.

Berman and Berman (1984) reached the same conclusion. They had first-year graduate students in social work, second-year graduate students in social work, and licensed social workers read a transcript of a clinical interview. The interviewee was a man with no serious problems who entered therapy to get more out of life. A cover sheet preceding the transcript described the interviewee as either "psychotic" or "normal." This manipulation followed Temerlin's (1968, 1970) methodology. Predictably, ratings of adjustment and prognosis were poorer when the interviewee was labeled psychotic than when he was labeled normal.

Additionally, ratings were negatively related to the subjects' level of training, that is, ratings of adjustment and prognosis became more negative as the amount of the subjects' training increased.

Berman and Berman concluded that psychiatric labels and clinical training are sources of bias in clinical judgment. They offered the following explanation for these effects. Clinicians undoubtedly learn negative stereotypes of mental illness at an early age like everyone else, stereotypes that are constantly reinforced by the media. Clinical training may reaffirm these stereotypes by emphasizing that mental illness can exist in the absence of overt symptoms and that clinicians must identify problem behaviors or weaknesses in order to be of help. With increased training clinicians become increasingly invested in the medical model and its notion that a diverse range of behaviors reflect underlying pathology. They may then become so skilled at recognizing pathology that almost any behavior becomes a "sign" of mental illness. Thus, clinicians with more training and experience may "find" more mental illness than less experienced clinicians, even in well-functioning individuals.

Berman and Berman made two recommendations to mitigate labeling bias. First, they suggested simply not using mental illness labels, or if this were not possible, at least withholding knowledge of previously assigned labels during the evaluation process. Second, they suggested that if educational programs in mental health (coursework and

continuing education) were to conceptualize problems within a social learning model rather than a disease model, "labeling bias effects might well be lessened and professionals' clinical judgments might not become more negative with increasing training and experience" (Berman & Berman, 1984, p. 41).

In summary, the Temerlin (1968, 1970), Langer and Abelson (1974), Caetano (1974), and Berman and Berman (1984) studies indicate that labeling bias can occur in clinical judgment, and that clinical experience can increase susceptibility to labeling bias. However, there exists little argument about the need to label psychopathology (Goodwin & Guze, 1984; Siassi, 1984). Diagnosis has long been considered the necessary beginning of successful treatment (Skodol & Spitzer, 1982) since few objective markers or specific etiologies have been identified for the major forms of mental illness (Wortis, 1982). This need for a common vocabulary of psychopathology led to the development of the Diagnostic and Statistical Manual of Mental Disorders, or DSM-I (American Psychiatric Association, 1952), which was followed by DSM-II (American Psychiatric Association, 1968), and now DSM-III (American Psychiatric Association, 1980).

Unfortunately, both DSM-I and DSM-II received considerable criticism from psychiatrists (Hines & Williams, 1975) and psychologists (Begelman, 1975), as neither yielded highly reliable classification schemes (Hersen & Barlow,

1976) and external validity was questionable as well (Eisler & Polak, 1971; Nathan, Zare, Simpson, & Ardberg, 1969). However, the multiaxial diagnostic system recently developed as DSM-III and adopted as the official guide of the American Psychiatric Association is generally regarded as a "major improvement in psychiatric diagnosis over previous classification systems" (Taylor, 1983, p. 13). DSM-III has received broad support across clinician groups (Light, 1982). Even behaviorists, who have in the past largely eschewed psychiatric diagnosis, have retrenched their position and are calling for clinicians to becomes familiar with and employ DSM-III (Goldstein & Hersen, 1984). In sum, rather than reject labels, the evidence suggests that clinicians and researchers should accept their necessity and study the process and impact of labeling.

The Labeling Theory of Mental Illness

Studies identifying labeling bias and clinical experience effects have been interpreted as evidence for the labeling theory of mental illness. Labeling theory, also known as societal reaction theory, was proposed primarily by sociologists as an alternative to the traditional psychiatric explanation of mental illness. Labeling theory has been called "the most dominant perspective among sociologists who study mental illness" (Gove, 1982, p. 320), and proponents of some version of labeling theory are also found in related

disciplines, including Szasz (1961, 1970) and Laing and Esterson (1964) in psychiatry, Goffman (1961) in anthropology, and Sarbin (1967, 1972) and Rosenhan (1973) in psychology. This section of the review of the literature summarizes the literature conceptualizing labeling theory and the research supporting and refuting the theory.

The history and development of labeling theory has been outlined by Gove (1980). Major early contributions to what would become labeling theory include Tannenbaum's (1938) statement concerning the development of criminal behavior and Lemert's (1951) book Social Pathology. Tannenbaum wrote

> The process of making the criminal is a process of tagging, defining, identifying, segregating, describing, emphasizing, making conscious and self-conscious; it becomes a way of stimulating, suggesting, emphasizing, and evoking the very traits that are complained of. (Tannenbaum, 1938, pp. 19-20)

Lemert outlined the fundamental distinction between primary and secondary deviance: Primary deviance is behavior that causes an individual to be labeled deviant, and secondary deviance (also called residual deviance) is behavior that is produced by placing an individual in a deviant role. According to Lemert,

> primary deviance is assumed to arise in a wide variety of social, cultural, and psychological contexts, and at best has only marginal implication for the psychic structure of the individual; it does not lead to symbolic reorganization at the level of self-regarding attitudes and social roles. Secondary deviation is deviant behavior or social roles based upon it, which becomes a means of defense, attack or adaptation to the overt and covert problems created by the societal reaction to primary deviation. (Lemert, 1967, p. 17)

These early contributions were followed by contributions from Garfinkel (1956), Goffman (1961), Erikson (1962), Kitsuse (1962), and Kitsuse and Cicourel (1963), and then landmark conceptualizations of labeling theory by Becker (1963) and Scheff (1966). Becker wrote

> Social groups create deviance by making rules whose infractions constitute deviance, and by applying those rules to particular people and labelling them as outsiders. From this point of view, deviance is not a quality of the act a person commits, but rather a consequence of the application by others of rules and sanctions to an "offender." The deviant is one to whom the label has successfully been applied; deviant behavior is behavior that people so label. (Becker, 1963, p. 9)

Scheff's work (Scheff, 1964, 1966, 1974, 1975, 1976, 1979) is commonly cited as the most explicit theoretical description of the labeling theory of mental illness. Researchers and theorists referring to labeling theory are usually referring to the work of Scheff. In his book Being Mentally Ill (Scheff, 1966) that formally outlined labeling theory, Scheff expanded the work of Becker by suggesting that "mental illness" could be conceptualized as the labeled violations of residual rules. He stated

> The culture of the group provides a vocabulary of terms for categorizing many norm violations: crime, perversion, drunkenness and bad manners are familiar examples. Each of these terms is derived from the type of norm broken, and ultimately from the type of behavior involved. After exhausting these categories, however, there is always a residue of the most diverse kinds of violations, for which the culture provides no explicit label. . . . The diverse kinds of rule-breaking for which our society provides no explicit label, and which, therefore, sometimes lead to the

> labeling of the violator as mentally ill, will be considered to be technically residual rule-breaking. (Scheff, 1966, pp. 33-34)

Scheff then outlined labeling theory as a set of nine propositions:

1. Residual rule-breaking arises from fundamentally diverse sources.
2. Relative to the rate of treated mental illness, the rate of unrecorded residual rule-breaking is extremely high.
3. Most residual rule-breaking is normalized and is of transitory significance.
4. Stereotyped imagery of mental disorder is learned in early childhood.
5. The stereotypes of insanity are continually reaffirmed, inadvertently, in ordinary social interaction.
6. Labeled deviants may be rewarded for playing the stereotyped deviant role.
7. Labeled deviants are punished when they attempt the return to conventional roles.
8. In the crisis occurring when a residual rule-breaker is publicly labeled the deviant is highly suggestible and may accept the proffered role of the insane as the only alternative.
9. Among residual rule-breakers, labeling is the single most important cause of careers of residual deviance. (Scheff, 1966, pp. 92-93)

In summary, according to Scheff's outline of labeling theory the concept of residual rule-breaking replaces the traditional psychiatric concept of mental illness (Horwitz, 1979). That is, psychiatric symptoms are viewed as labeled violations of social norms rather than actual intrapsychic disturbances. Furthermore, labeling is seen as beginning a self-fulfilling prophecy which can result in permanent acceptance of the role of mentally ill.

The studies summarized in the previous section of this review of the literature have been interpreted as support

for labeling theory. For example, Caetano (1974) presented his manipulation of clinical experience and the label "mental patient" as an experimental paradigm constructed specifically to test labeling theory. Caetano hypothesized "if labeling theory is valid, both clinical experience and the suggestion of mental illness should account for a significant amount of the variation in diagnostic scores" (Caetano, 1974, p. 256). Caetano reported that his results partially confirmed this hypothesis, and that psychiatry's assumption that patients' behaviors are the primary determinants of diagnoses was successfully challenged by such results.

The other studies previously summarized that identify labeling bias and clinical experience effects have been similarly interpreted. Scheff (1974) conducted a review of these and all other studies that appeared relevant to labeling theory and concluded that 13 of 18 studies provisionally supported labeling theory. According to Scheff, the balance of the evidence in 1974 favored the labeling perspective over the psychiatric perspective of mental illness. In a second review, Krohn and Akers (1977) assessed the results of 26 studies conducted on the admission and discharge of mental patients and reached conclusions similar to Scheff's. Krohn and Akers found that 23 studies generally supported labeling theory, and only three did not. In concert with labeling theory, both Scheff

and Krohn and Akers emphasized in their reviews the importance of social and contextual rather than psychiatric factors in the development of mental illness. In a third review, Conover (1976) concurred with Scheff and Krohn and Akers that the available evidence supported labeling theory as "a viable guide to research and practice" (Conover, 1976, p. 243).

Labeling theory has enjoyed considerable popularity, particularly among sociologists studying mental illness. Since the mid-1960s, however, an increasing number of researchers have marshalled evidence discounting labeling theory. The reviews of Scheff (1974), Krohn and Akers (1977), and Conover (1976) supporting labeling theory have become a distinct minority. At least 28 studies offered unfavorable critiques of labeling theory through 1976 (Conover, 1976), and this level of criticism has continued (Scull, 1984). Unfavorable critiques by Gibbs (1972), Davis (1972), Hirschi (1973), Manning (1973), and especially Gove (1970a, 1970b, 1975, 1979, 1980, 1982) have demonstrated the inability of labeling theory to adequately explain mental illness. Gove's critiques will be considered in detail here.

Gove stated his initial understanding of labeling theory as follows:

> The argument of the societal reaction theorists is that persons who have passed through a degradation ceremony and have been forced to become members of a deviant group have experienced a profound and frequently irreversible socialization process. They have acquired an inferiority status and have

> developed a deviant world view and the knowledge and skill that go with it. And perhaps equally important, they have developed a deviant self-image based upon the image of themselves they receive through the actions of others. Although the societal reaction perspective of deviance has been very much in vogue during the 1960s, most of the work based on this perspective has been intuitive and/or theoretical, and there has been very little systematic evaluation and testing of the perspective. (Gove, 1970a, p. 875)

Gove (1975) then disagreed with Scheff's (1974) interpretation of the relevant research as supporting labeling theory. He reviewed the same 18 studies that Scheff had reviewed and found "the evidence in support of labelling theory to be quite weak" (Gove, 1975, p. 247). He continued "it is very clear that if we had to choose between the psychiatric perspective and the labeling perspective, the evidence overwhelmingly supports the psychiatric perspective" (Gove, 1975, p. 247). Gove concluded that labeling theory's explanation of mental illness was incorrect and that the evidence indicated that most psychiatric hospitalizations are of individuals suffering from real and serious disturbances.

What developed has become known as the Scheff-Gove debate (Imershein & Simons, 1976). Following Gove's (1970a) initial criticism of Scheff's (1966) outline of labeling theory, a series of rebuttals and replies to rebuttals ensued (see Horwitz, 1979). Through it all the central issue remained whether the labeling perspective (Scheff) of mental illness or the psychiatric perspective (Gove) was the

more powerful (Horwitz, 1979). Although as mentioned some reviewers have supported Scheff (1964, 1966, 1974, 1975, 1976, 1979), the weight of the evidence presented (Gove, 1970a, 1970b, 1975, 1979) and updated (Gove, 1980, 1982) by Gove convincingly reveals the untenable empirical base which labeling theory rests upon. The remainder of this section of the review of the literature summarizes this evidence refuting labeling theory.

According to Gove, labeling theorists continue to support an inaccurate theoretical position because they have ignored decades of advances in the very discipline, psychiatry, that treats the mentally ill. Gove argues that "in the past 25 years there have been so many developments in psychiatry that the issues raised by labelling theory have been largely resolved and that from a pragmatic perspective a general labelling explanation of mental illness is no longer tenable" (Gove, 1982, p. 307).

Gove describes five major areas of change in the practice of psychiatry that refute the labeling perspective of mental illness: the rate and place of treatment, patients' civil rights, attitudes toward the mentally ill, the relationship between individual characteristics and mental illness, and the effectiveness of treatment. These five areas of change are summarized below.

Labeling theory asserts that mental patients become socialized into the role of chronically ill. However, changes in the rate and place of psychiatric treatment

counter this assertion. Examination of treatment statistics indicates that the resident population of individuals in public mental hospitals has declined every year since 1955. This decline is the result of a dramatic decrease in the length of hospital stays, not a decrease in the number of admissions (Klerman & Schechter, 1981). There has also been a dramatic shift in the place of treatment, away from public mental hospitals and toward community treatment centers (Regier, Goldberg, & Taube, 1978). Thus, whereas in 1955 most individuals requiring psychiatric treatment received relatively long-term stays in public mental hospitals, the vast majority of treatment now occurs in outpatient settings and inpatient stays, when necessary, are much briefer (Gove, 1982). Contrary to labeling theory, these changes make it less likely that individuals receiving psychiatric treatment are becoming permanently socialized into the role of mental patient.

Labeling theorists have argued that individuals' civil rights are violated during the commitment process. However, substantial progress on civil rights issues has quieted this complaint. Key legal decisions in the last 10 to 15 years have held that involuntarily committed individuals have constitutional rights to individual treatment aimed at improving their condition and that "a state cannot constitutionally confine a nondangerous individual who is capable of surviving safely in freedom by himself or with the help of willing and responsible family members or friends" (Crane,

Zonana, & Wiser, 1977, p. 827). As a result of these and other decisions, hospitals have set up review boards to protect involuntary patients, and there is some evidence that hospitals' psychiatric care has improved (Gove & Fain, 1977; Kaufman, 1979). Individuals can now be committed only if there is clear and convincing evidence that they are both mentally ill and dangerous to themselves or others (Feigelson, Davis, Mackinnon, Shands, & Schwartz, 1978; Morrissey, Tessler, & Farrin, 1979). In sum, the abuse of patients' civil rights is certainly less an issue now than when labeling theory was developed (Gove, 1982).

Negative attitudes toward the mentally ill created by the stigmatizing effects of psychiatric labels (Rabkin, 1974) are often presented as among the strongest evidence for labeling theory. However, it appears that individuals requiring psychiatric treatment are increasingly being thought of as suffering a "nervous breakdown" rather than labeled as mentally ill (Rabkin, 1979). Review of the literature on stigma (Gove, 1982) also indicates that

1. Although being labeled mentally ill may be stigmatizing, being labeled a former mental patient is not (Olmstead & Durham, 1976).

2. Persons who have had experience with the mentally ill are less rejecting than unexperienced persons (Trute & Loewen, 1978).

3. Family members tend not to be rejecting of the mentally ill, as close ties override stigma (Kriesman & Joy, 1974).

4. Although employers are apt to express prejudice against the mentally ill, they do not discriminate against former mental patients in their hiring practices (Huffine & Clausen, 1979; Olshansky, Brob, & Dahl, 1960).

5. Despite the stigmatizing effects of labels, the great majority of hospitalized mental patients have favorable attitudes toward the hospital, their treatment, and the staff (Weinstein, 1982).

Labeling theorists claim that there are no differences between individuals labeled mentally ill and individuals not so labeled, and that individuals in lower classes are more likely to be labeled mentally ill solely due to discrimination. However, the following relationships between individual characteristics and mental illness (Gove, 1982) have received strong empirical support:

1. There is a genetic component to schizophrenic and affective disorders (Gershon, Bunney, & Leckman, 1976).

2. Individuals who experience a number of changes in their lives are likely to experience emotional difficulties (Myers, Lindenthal, & Pepper, 1975).

3. Individuals in lower classes experience more emotional difficulties (Kulka, Veroff, & Donovan, 1979), and may therefore be more likely to become mentally ill and require psychiatric treatment.

Finally, the increased effectiveness of psychiatric treatment challenges the labeling perspective of mental illness. The efficacy of psychopharmacological treatment, particularly with tranquilizers and antidepressants, is well established (Berger, 1978). Reviews of psychotherapy outcome studies (Bergin & Lambert, 1980) add that psychotherapy generally appears to be a productive and helpful treatment or adjunct to treatment for various disorders. An increase in the reliability of psychiatric diagnosis has been provided by the development of DSM-III (American Psychiatric Association, 1980). The necessity of continued treatment has been highlighted by the deinstitutionalization movement, as former inpatients have clearly been helped by appropriate maintenance therapy (Hansell, 1978). In short, labeling theory's claim that psychiatric treatment is irrelevant or broadly ineffective is not supported by the literature.

In summary, according to Gove (1982) labeling theory continues to be the dominant perspective among sociologists who study mental illness because decades of advances in the practice of psychiatry have been ignored. Gove claims that "for all practical purposes the labeling explanation of mental illness is of historical interest only" (Gove, 1979, p. 301). However, in each of his critiques of labeling theory (1970a, 1970b, 1975, 1979, 1980, 1982) Gove is careful to emphasize that labeling theory does point to

real, important, and potentially debilitating processes. He strongly encourages researchers to integrate these processes into the psychiatric perspective of mental illness (Gove, 1979). The present study suggests that labeling bias and clinical experience effects can best be integrated into the psychiatric perspective by considering them to be cognitive errors on the part of clinicians.

Inaccuracy in Clinical Judgment

Deinhardt (1983) emphasizes that "clinical judgment" has a dual meaning: "It can mean the final product of clinical thinking and denote a conclusion, interpretation, or judgment. It can also mean the process of clinical thinking, that is, the tacit methodology the clinician uses to produce judgments" (Deinhardt, 1983, p. 49). Other researchers have similarly conceptualized clinical judgment as both a subjective process of inference and the outcome of such a process, that is, a conclusion and a prediction (Arkes, 1981; Einhorn & Hogarth, 1978; Heider, 1980; Holt, 1961; Lopez, 1984; Meehl, 1954). Numerous judgment processes have been identified by researchers that contribute to erroneous outcomes of clinical judgments. This section of the review of the literature summarizes the research investigating inaccuracy in clinical judgment.

Reed and Jackson (1975) define accuracy in clinical judgment as

> a person's ability, given limited information about a target person (patient), to judge correctly other pertinent characteristics about that person and to identify behavioral exemplars as part of a pattern of behavioral consistencies. (Reed & Jackson, 1975, p. 475)

That clinicians can err in making judgments about a person's psychological status has been noted since the early part of the century (Franz, 1920). Perhaps the most compelling criticism of clinicians' accuracy has been provided by what has become known as the clinical versus statistical prediction comparison. Clinicians' judgments, or predictions, have generally been considered subjective events and compared with objective, statistical predictions. The common image is of a clinician making predictions based on experience and clinical lore and a statistician constructing actuarial models for prediction (Holt, 1961). The clinical versus statistical prediction comparison began in earnest in the 1950s in large part due to Meehl's influential work (Meehl, 1954, 1956, 1957, 1959, 1960). In his classic monograph Clinical versus Statistical Prediction Meehl (1954) reviewed the empirical literature comparing clinical and statistical prediction methods and concluded that the reliability of clinical prediction was in fact less than that obtained with simple statistical procedures. Although he emphasized the necessity of retaining human judgment in many circumstances, Meehl concluded that clinical psychologists could not predict the outcome of treatment better than simple regression formulas combining available diagnostic information.

Meehl's monograph received considerable attention, and researchers began to report similar findings that the reliability of clinicians' judgments was low (Goldberg & Werts, 1966; Little & Schneidman, 1959; Phelan, 1964; Wallach & Schoof, 1965). Reviewers of the substantial body of research that has accumulated since Meehl's monograph was published have generally concurred with Meehl that statistical prediction reliably surpasses clinical prediction (Dawes, 1976; Goldberg, 1968; Sawyer, 1966; Wiggins, 1973, 1981). Although supporters of the clinician have contended that the comparison is invalid because clinical judgment cannot be objectified (Holt, 1961; Richards, 1963), the weight of the evidence falls heavily in favor of statistical and actuarial models of prediction.

Stimulated by the clinical versus statistical prediction comparison, researchers began to focus on the effects of professional training on accuracy, the relationship between the amount of information available to a judge and his accuracy, the generality or specificity of judgmental accuracy, and the characteristics of "good" judges (Mischel, 1968). Reviews of these four lines of research uniformly reach pessimistic conclusions regarding the general accuracy of clinicians' judgments. These reviews are summarized below.

Goldberg (1968) reviewed 10 studies investigating the effects of training on accuracy and concluded that all 10 reached the surprising and discouraging finding that

the amount of professional training and experience a judge possesses is not related to his judgmental accuracy. Goldberg also reviewed another 14 studies that demonstrated the "equally disheartening" (Goldberg, 1968, p. 484) conclusion that the amount of information available to a judge is not related to the accuracy of his judgments, that is, that increasing the amount of information presented to a judge does not result in an increase in accuracy. According to Wiggins' (1973) review of the literature investigating the possible generality of judgmental accuracy, generality of accuracy across persons and situations is doubtful (see also Mischel, 1968). Finally, although some characteristics of "good" judges have been tentatively identified--intelligence, esthetic interests, self-insight, emotional adjustment, and social skill--these are not necessarily traits which clinicians inherently possess while other persons do not (Taft, 1955).

In sum, although it has been frequently assumed that clinicians are generally more accurate in their judgments than laymen due to training, experience, and the ability to integrate large amounts of data, the evidence indicates that "claims of superior judging ability for clinicians must be viewed with considerable skepticism" (Wiggins, 1973, p. 180). Recent reviews (Cline, 1985; Deinhardt, 1983; Levin, 1984; Patterson, 1983) reaffirm this skepticism. Few researchers would argue with Reed and Jackson that "the attempt to demonstrate accuracy of judgment has been

abandoned" (Reed & Jackson, 1975, p. 475). Researchers have turned instead toward investigations of sources of inaccuracy in the judgment process.

Four reviews, by Lourens (1979), Arkes (1981), Einhorn and Hogarth (1978), and Kruglanski and Ajzen (1983), have attempted to outline the range of impediments to accuracy in clinical judgment that have been identified. These reviews are summarized below.

Lourens (1979) presents a two-stage model of clinical judgment in which clinicians sift and then interpret evidence, with the possibility of inferential errors occurring in both stages. In the first stage of Lourens' model, the following errors occur when a clinician sifts evidence: (a) A significant portion of information is lost as transmission from input to output (judgment) is incomplete (Bieri, Atkins, Briar, Leaman, Miller, & Tripodi, 1966; Reed & Jackson, 1975); (b) cues receive differential weighing, that is, judgments as to how to weigh cues are often invalid (Chapman & Chapman, 1969; Lueger & Petzel, 1979); (c) differentiation of cues is inadequate, leading to anchoring errors (a shift in judgment caused by attention to end stimuli), assimilation errors (the underestimation of the difference between two stimuli), or contrast errors (the tendency to exaggerate the distance between two stimuli) (Bieri, Atkins, Briar, Leaman, Miller, & Tripodi, 1966); and (d) evidence is inconsistently filtered or conceptualized in differing theoretical terms (Anderson, 1970).

In the second stage of Lourens' model, the following errors occur when a clinician interprets evidence: (a) Evidence is contaminated by memory and the clinician's past experiences (Helson, 1964; Slovic & Lichtenstein, 1971); (b) interactions between cues influence judgments (Hoffman, Slovic, & Rorer, 1968; Goldberg, 1968); (c) probabilities of conjunctive events' occurring are overestimated, and probabilities of disjunctive events' occurring are underestimated (Tversky & Kahneman, 1974); (d) invalid hunches or presuppositions are applied to the judgment process (De Rivera, 1968); (e) estimations of probabilities are constrained by limitations of the clinician's setting, the influence of prior experiences with other clientele, and failure to account for regression toward the mean (Kahneman & Tversky, 1973); and (f) reliance on judgmental heuristics leads to biased probability estimates (Tversky & Kahneman, 1974).

Arkes (1981) summarizes the limitations to accuracy in clinical judgment in his review as (a) the inability to assess the covariation of symptoms (e.g., delusions) and outcomes (e.g., psychotic episode) accurately (Nisbett & Ross, 1980; Smedslund, 1966); (b) the influence of preconceived notions or expectancies, that is, illusory correlations (Chapman & Chapman, 1969; Kurtz & Garfield, 1978); (c) negligible awareness of the factors that influence clinical judgment (Oskamp, 1967; Summers, Taliaferro, & Fletcher, 1969); (d) serious overconfidence that

diagnosticians place in their diagnoses (Einhorn & Hogarth, 1978; Holsopple & Phelan, 1954; Oskamp, 1965); and (e) hindsight bias, that is, finding evidence in a rich source of data to support almost any diagnosis (Arkes, Wortmann, Saville, & Harkness, 1981; Fischoff, 1975).

Einhorn and Hogarth's (1978) review identifies the following limitations to accurate clinical judgment: (a) failure to recognize the statistical notions of randomness, variance, and sampling variability (Hogarth, 1975); (b) an inability to revise incorrect opinions (Edwards, 1968); (c) reliance on judgmental heuristics, leading to biased probability estimates (Tversky & Kahneman, 1974); (d) poor calibration of judgments of probabilities with their actual empirical frequencies (Lichtenstein, Fischoff, & Phillips, 1977); (e) unavailability of outcome feedback (Castellan, 1977); (f) the ignoring of base rate information (Lyon & Slovic, 1976; Nisbett, Borgida, Crandall, & Reed, 1976); and (g) hindsight bias (Fischoff, 1975).

Kruglanski and Ajzen's (1983) review notes that the psychological literature classifies biases in human judgment as either motivational or cognitive in origin (Ross, 1977). Motivational biases are characterized by a tendency to form and hold beliefs that serve the individual's needs and desires. They include (a) a tendency to attribute successes to internal factors and to blame external factors for failures (Beckman, 1970; Johnson, Feigenbaum & Weiby, 1964); (b) the desire to exercise control, leading individuals to

attribute events to controllable factors rather than to factors over which they have no control (Kelley, 1971); (c) the desire of individuals to control the outcomes of others' behaviors when such behaviors have consequences for themselves (Jones & Davis, 1965); (d) the need for people to believe that the world is a just and fair place (Lerner, 1970); and (e) the need to believe in the avoidability of accidental misfortunes and thus in the ability to control one's own fate (Walster, 1966).

Cognitive biases, according to Kruglanski and Ajzen's review, arise because people have a limited ability to properly process all the information available to them. They include (a) erroneously concluding that a sample of information is representative of the population as a whole (Nisbett & Ross, 1980); (b) selectively focusing on only those features of a situation that are perceptually salient (Jones & Nisbett, 1971); (c) reliance on the availability heuristic, that is, assessing the probability of an event by the ease with which similar instances can be brought to mind (Tversky & Kahneman, 1974); (d) erroneously perceiving a covariation of traits (Asch, 1946); (e) reliance on the representative heuristic, that is, appraising the likelihood that object A belongs to class B by the degree to which A is representative of (resembles) B (Kahneman & Tversky, 1973); and (f) the forming of an initial hypothesis, or anchor, that biases interpretation of new information toward the

initial values (Ajzen, Dalto, & Blyth, 1979; Tversky & Kahneman, 1974).

In addition to reviewing the range of impediments to accuracy in clinical judgment, Lourens (1979), Arkes (1981), Einhorn and Hogarth (1978), and Kruglanski and Ajzen (1983) make specific suggestions on how to improve the clinical judgment process.

Lourens (1979) recommends (a) utilizing instrumentation to assure an independent check of perceptions against reality (De Rivera, 1968); (b) training clinicians in statistical decision theory (Goldberg, 1968); (c) selecting clinicians less susceptible to illusory correlation and more receptive to training (Golding & Rorer, 1972); and (d) placing more reliance on actuarial procedures (Goldberg, 1970).

Arkes recommends the following as "debiasing techniques" (Arkes, 1981, p. 336): (a) forcing clinicians to consider alternatives to reduce unwarranted overconfidence (Slovic & Fischoff, 1977); (b) presenting clinicians with a tutorial in Bayesian statistics (Galen & Gambino, 1975); and (c) decreasing clinicians' reliance on memory (Arkes & Harkness, 1980).

Einhorn and Hogarth (1978) recommend that clinicians improve accuracy by improving their ability to learn from experience. They suggested this could be accomplished by: (a) formal training in experimental design (Einhorn, 1972); (b) teaching clinicians the logic of control groups and

baseline predictions (Einhorn, 1972); (c) employing a model of the environment to increase awareness of environmental effects on outcomes (Hammond, 1971); and (d) simply keeping a box score of accurate and inaccurate judgments (Goldberg, 1968).

Kruglanski and Ajzen (1983) noted in their review that judgmental biases are traditionally classified in the psychological literature as either motivational or cognitive. According to Kruglanski and Ajzen, the literature recommends that motivational biases can be reduced or eliminated by any process which reduces or eliminates the need to bring judgments in line with one's needs and desires, and cognitive biases can be reduced or eliminated by any process which decreases intuitive judgments and increases more objective modes of information processing.

Other researchers have also added recommendations for improving the accuracy of the clinical judgment process. For example, Strasburger and Jackson (1977) concluded that accuracy could be increased with increased practice and feedback, and Schlesinger (1984) theorized that self-analytic exploration can positively affect clinicians' judgment ability. However, few if any recommendations for improving accuracy in clinical judgment have been widely implemented, and clinicians' judgments of diagnosis, prognosis, and so on continue to be viewed as subjective processes susceptible to systematic errors and biases (Levin, 1984; Patterson, 1983).

This section of the review of the literature has summarized the literature identifying numerous impediments to accuracy in clinical judgment. Researchers have focused primarily on clinicians' erroneous cognitive processes. The weight of this evidence suggests interpreting the labeling bias and clinical experience effects reviewed earlier as cognitive errors on the part of clinicians. Such an interpretation appears both more accurate and more useful than that provided by the labeling theory of mental illness.

Anchoring Effects in Clinical Judgment

As evidenced by the reviews of Arkes (1981), Einhorn and Hogarth (1978), Kruglanski and Ajzen (1983), and Lourens (1979) listing numerous impediments to accuracy in clinical judgment, the old notion that people generally process information in a reasonable fashion and make judgments in accordance with the normative principles of statistical models has undergone drastic revision. Kruglanski and Ajzen conclude in their review of the literature that "people's judgments are no longer viewed as following rationally and impartially from the relevant information. Instead, attributions and predictions are said to be subject to systematic biases and errors" (Kruglanski & Ajzen, 1983, p. 2). Kruglanski and Ajzen further conclude that "dissatisfaction with the normative models as descriptions of the inference process can perhaps be traced to the work of Tversky and Kahneman" (Kruglanski & Ajzen, 1983, p. 2). The prominence

of Tversky and Kahneman's theory of judgmental heuristics was evident in the previous section of this review of the literature. Ross (1977) adds "any adequate portrayal of the intuitive psychologist and his shortcomings must bring these judgmental heuristics to center stage and assign them a leading role" (Ross, 1977, p. 388). This section of the review of the literature summarizes the research investigating anchoring as a judgmental heuristic (Tversky & Kahneman, 1974) that can bias clinical judgment.

Anchoring effects were originally studied only in judgments of sensory stimuli. This early psychophysical research identified two characteristic anchoring phenomena, contrast and assimilation. Contrast is a tendency to shift judgments away from an anchor stimulus, and assimilation is a tendency to shift judgments toward an anchor stimulus. According to Bieri, Atkins, Briar, Leaman, Miller, and Tripodi (1966), both effects were considered "distortions of perception which affect one's judgment of stimuli" (Bieri et al., 1966, p. 154). A variety of simple early theories and models were developed to explain these two anchoring effects. These included Johnson's (1944) generalization theory, Helson's (1947) adaptation-level theory, Volkmann's (1951) "rubberband" theory, Peak's (1958) activation surface model, and Sherif and Hovland's (1961) assimilation and contrast model. However, none of these early psychophysical theories or models won universal acceptance (Bieri, Orcutt, & Leaman, 1963; Campbell, Hunt, & Lewis, 1957; Levy, 1961).

Classic studies by Meehl (1960), Oskamp (1965), and Bieri, Orcutt, and Leaman (1963) then suggested that anchoring effects also occur in judgments of clinical stimuli. Meehl found that clinicians' ratings (Q-sorts) of clients stabilized after a very short time, only two to four psychotherapy sessions, and Oskamp found that the accuracy of clinicians' judgments of a one-paragraph case study did not increase with any amount of increased information. Bieri, Orcutt, and Leaman were among the first researchers to actually assess and report anchoring effects in clinical judgment. It soon became widely accepted (Bieri et al., 1966) that an initial hypothesis or judgment serves as an anchor that guides clinicians' interpretation of subsequent information. The early psychophysical theories and models of anchoring were generalized to complex clinical judgment situations, and new theories began to be developed as well.

The most prominent theoretical conceptualization of anchoring effects that has been developed is that of Tversky and Kahneman (Kahneman & Tversky, 1972, 1973; Tversky & Kahneman, 1971, 1973, 1974). Tversky and Kahneman identified representativeness, availability, and anchoring as heuristics, or problem-solving techniques, that can lead to systematic inferential biases. Tversky and Kahneman were interested in investigating judgments people make under uncertain conditions, e.g., the outcome of an election, the

guilt of a defendant, the future value of the dollar, or the diagnosis of a patient. They concluded that

> people rely on a limited number of heuristic principles which reduce the complex tasks of assigning probabilities and predicting values to simpler judgmental operations. In general, these heuristics are quite useful, but sometimes they lead to severe and systematic errors. (Tversky & Kahneman, 1974, p. 1124)

Representativeness occurs when probabilities are evaluated by the degree to which A is representative of B, that is, the degree to which A resembles B. The inferential biases that can result from use of the representativeness heuristic were summarized by Tversky and Kahneman (1973) as insensitivity to prior probability of outcomes, insensitivity to sample size, misconceptions of chance, insensitivity to predictability, the illusion of validity, and misconceptions of regression.

Availability occurs when people assess the probability of an event by the ease with which instances or occurrences can be brought to mind. The inferential biases that can result from use of the availability heuristic were summarized by Tversky and Kahneman (1973) as biases due to the retrievability of instances, biases due to the effectiveness of a search set, biases of imaginability, and illusory correlation.

Anchoring occurs when people make initial judgments and insufficiently adjust these judgments as additional information is processed to yield a biased final answer. The inferential biases that can result from use of the anchoring

heuristic were summarized by Tversky and Kahneman (1973) as insufficient adjustment, biases in the evaluation of conjunctive and disjunctive events, and biases in the assessment of subjective probability distribution.

The stimulus which provides the initial judgment or starting point is called the "anchor." To demonstrate the existence of anchors, Tversky and Kahneman (1973) asked subjects to estimate quantities (e.g., the percentage of countries in the United Nations from the African continent). For each quantity, a number between 0 and 100 was selected as a starting point by spinning a wheel of fortune in the presence of the subjects. Subjects were then instructed to estimate whether the actual quantity was higher or lower than this randomly selected number (the anchor) and to estimate the exact value of the quantity. Results indicated that the randomly selected number had a marked effect on estimates, and payoffs for accuracy did not reduce this anchoring effect. Tversky and Kahneman concluded "different starting points yield different estimates, which are biased toward the initial values. We call this phenomenon anchoring" (Tversky & Kahneman, 1974, p. 1128).

On the basis of Tversky and Kahneman's conceptualization of anchoring, Friedlander and Stockman (1983) hypothesized that clinicians' estimates of pathology and prognosis would be a function of the timing of their exposure to pathological case material. That is, they predicted that clinicians exposed to pathological material (the anchor)

appearing early in interview notes would make a lower estimate of a client than clinicians exposed to the same information appearing later in the same interview notes.

Results supported this hypothesis for clinicians' ratings of a moderately disturbed client, but not a severely disturbed client. For the moderately disturbed client, significant differences were found among judgments of clinicians exposed to pathological case material appearing early vs. late in interview notes. Friedlander and Stockman concluded "an experienced sample of clinicians demonstrated significant anchoring in their final pathology and prognosis estimates of a moderately disturbed client" (Friedlander & Stockman, 1983, p. 642). For the severely disturbed client, Friedlander and Stockman concluded that clinicians may tend to disregard severe pathological material about a client viewed initially as less disturbed.

Friedlander and Phillips (1984) attempted to replicate Friedlander and Stockman's anchoring effect, with undergraduate students serving as subjects. Their methodology was identical to that Friedlander and Stockman had employed to identify anchoring in a moderately disturbed client. Friedlander and Phillips hypothesized that anchoring would occur, but that it could be prevented by presenting debiasing information to the students prior to the judgment task. They reasoned "judges could be debiased by being warned of anchoring errors and how to avoid them" (Friedlander &

Phillips, 1984, p. 367). However, results indicated that debiasing was irrelevant, as Friedlander and Stockman's anchoring effect was not replicated: Students were not susceptible to the anchoring manipulation and therefore were not in need of debiasing. Additionally, clinicians had expressed more confidence in their judgments in the Friedlander and Stockman study than did students in Friedlander and Phillips' study. Friedlander and Phillips concluded that these unexpected results suggest that clinicians may be more susceptible to anchoring effects than students or inexperienced judges.

The present study suggests that the Friedlander and Stockman (1983) and Friedlander and Phillips (1984) studies describe the same cognitive processes Temerlin (1968), Langer and Abelson (1974), Caetano (1974), and Berman and Berman (1984) described in their investigations of labeling bias and clinical experience effects. An anchoring explanation fits the data in all these studies: Assigning labels suggesting pathology creates an anchor that biases judgments of subsequent information, and experience with labels increases susceptibility to this anchoring effect. The present study concludes that Tversky and Kahneman's conceptualization of anchoring provides a more accurate and useful explanation for labeling bias and clinical experience effects than that provided by labeling theory. Rather than conclude that mental illness is a myth, it is apparent that

clinicians are prone to certain inferential errors and biases just as laymen are.

Support for interpreting labeling bias and clinical experience effects as anchoring phenomena is provided by research on impression formation and maintenance. Mischel (1968) reviewed the early impression formation research (Anderson, 1965; Asch, 1946; Bruner & Tagiuri, 1954; Dailey, 1952; Tagiuri & Petrullo, 1958; Wishner, 1960) and concluded "interpersonal stereotypes are developed rapidly from minimal information and constrain the way in which later new data are interpreted" (Mischel, 1968, p. 127). Other early researchers added that once a perceiver categorizes a person he tends to use new data to confirm his categorization rather than to generate new hypotheses (Rubin & Shontz, 1960; Sines, 1959). The more recent research on impression formation and maintenance also mimics anchoring effects: Even when an initial hypothesis is formed that is merely tentative, it is still likely to exert a biasing influence (Arkes, 1981; Fiske, 1976; Ross, Lepper, Strack, & Steinmetz, 1977). In fact, even when faced with complete negation of their original beliefs, judges will often continue to maintain their original impressions (Ross, Lepper, & Hubbard, 1975).

Summary

The first section of the review of the literature summarized the research investigating labeling bias in

clinical judgment. It was concluded that although labels can bias clinical judgments and clinical experience can increase susceptibility to labeling bias, these effects do not warrant dismissing classification systems of mental disorders.

The second section of the review outlined the labeling theory of mental illness and summarized the literature supporting and refuting labeling theory. The labeling bias and clinical experience effects summarized in the first section have been interpreted as supporting labeling theory. However, it was concluded that the evidence refuting labeling theory is overwhelming, and alternative explanations for the important processes identified by labeling theorists are needed.

The third section of the review summarized the research investigating inaccuracy in clinical judgment. This research has identified a wide range of impediments to accurate clinical judgments, emphasizing the fallibility of clinicians' cognitive processes. It was concluded that interpreting labeling bias and clinical experience effects as cognitive errors on the part of clinicians is more accurate and useful than viewing them as support for labeling theory.

The fourth section of the review summarized the research investigating anchoring as a judgmental heuristic that can bias clinical judgment. The anchoring heuristic was selected from the research on inaccuracy in clinical

judgment as the best explanation for labeling bias and clinical experience effects. An anchoring interpretation was applied to the previous studies identifying labeling bias and clinical experience effects, and support for this interpretation was provided by related research on impression formation and maintenance.

CHAPTER III

METHODOLOGY

Subjects

There were two groups of experimental subjects in the present study: clinicians and students. The students were sampled from an undergraduate abnormal psychology class at the University of Florida. Students enrolled in abnormal psychology for the Fall 1986 semester were invited to participate and offered extra credit for participation. Students were ruled ineligible if they reported any prior clinical experience. The clinicians were sampled from Division 29 (Psychotherapy) of the American Psychological Association. Clinicians listed alphabetically in the 1985 Directory of the American Psychological Association as Fellows or Members of Division 29 were eligible for participation. From this list a total N of 400 clinicians were selected using a systematic random sampling plan. The names of the clinicians were numbered serially from one to N. A sampling interval, k, was determined by N/400. Starting from a point in the first interval selected from a table of random numbers, every kth clinician was then selected until N=400 clinicians were selected. Each of

these 400 clinicians was mailed the experimental materials and invited to participate.

Materials

The experimental materials consisted of a cover sheet (for clinicians Appendix A, for students Appendix B), a case study (Appendix C, D, E, or F), the Global Assessment Scale (Appendix G), and the Prognostic Scale (Appendix H). These experimental materials are described in turn below.

Cover Sheet for Clinicians

The cover sheet mailed to 400 clinicians (Appendix A) identified the study as a doctoral dissertation and introduced it as an investigation of "factors influencing clinical judgment" (Friedlander & Stockman, 1983, p. 639). This was the deliberately ambiguous introduction utilized by Friedlander and Stockman to facilitate the subtle assessment of inferential biases in clinical judgment. The cover sheet invited clinicians to participate, stressed the importance of their cooperation, and emphasized that participation would only require approximately 10 minutes. A paragraph was included that served as a declaration of informed consent. Clinicians were assured that they could contact the experimenter with questions, and appreciation was expressed for their possible participation. The cover sheet ended by asking clinicians, should they agree to participate,

to provide their age, sex, and years of experience as demographic information.

Cover Sheet for Students

The cover sheet administered to students (Appendix B) varied minimally from the cover sheet for clinicians. The study was again identified as a doctoral dissertation and introduced as an investigation of "factors influencing clinical judgment." Students were asked to complete judgment tasks and the importance of their cooperation was stressed. The same informed consent paragraph was included. Students were assured they could ask any questions they might have, and appreciation was expressed for their participation in the hope of maximizing motivation to participate accurately. The cover sheet ended by asking students to provide their sex, class standing, declared major, and number of credit hours of coursework in psychology as demographic information.

Case Study

The case study was chosen from Spitzer, Skodol, Gibbon, and Williams' Psychopathology: A Case Book (1983). Spitzer et al. published Psychopathology: A Case Book as a collection of case studies they found useful in teaching DSM-III to clinicians. They discovered that studying relatively brief case studies of real patients, edited to

focus on information relevant to differential diagnosis, was the most effective way for clinicians to gain experience with DSM-III diagnoses. Spitzer et al. emphasize that each case study was prepared to ensure that all information relevant to making a diagnosis was included. According to Spitzer et al., each case study represents as accurately as possible an example of a DSM-III diagnostic category.

These case studies conform to Heverly, Fitt, and Newman's (1984) model for effective analogue research stimuli. Heverly et al. noted that the analogue research literature lists several disadvantages in relying on live or taped clinical interviews (Kushner, 1978; Kushner, Bordin, & Ryan, 1979) and therefore focused on constructing a model for effective written case vignettes. According to their model, written case vignettes are ideally suited for analogue research if they (a) summarize the relevant features of actual case histories, and (b) differ from each other only regarding the clinical factors being studied. Heverly et al. conclude that written case vignettes constructed using their model "are most useful when the judgments being made about individuals are susceptible to bias or distortion" (Heverly, Fitt, & Newman, 1984, p. 53). Kopta, Newman, McGovern, and Sandrock (1986) recently concurred that written case vignettes are preferable to live or videotaped interviews when examining cognitive processes. Kopta et al. noted that a review of the empirical literature

"indicates that to sample an underlying cognitive process, a procedure should minimize interference from extraneous variables not under experimental control. This favors written vignettes for this type of study" (Kopta, Newman, McGovern, and Sandrock, 1986, p. 373).

The case studies developed by Spitzer et al. depart from the model constructed by Heverly et al. only in that they are longer than Heverly et al. recommend as necessary. Researchers investigating judgmental bias have concluded that greater bias is found by studies that present little information about a stimulus person than studies where much more information is presented (Nieva & Gutek, 1980). Apparently subjects tend to rely more on stereotypes when they are asked to make judgments from minimal amounts of information (Mischel, 1968). Thus, utilizing the case studies developed by Spitzer et al. should provide a more stringent test of bias in clinical judgment than would shorter case vignettes such as Heverly et al. recommend. Longer case studies were also deemed necessary in order to examine possible anchoring effects.

Spitzer et al. recommend that their case studies be used to (a) help experienced as well as inexperienced clinicians learn the new DSM-III concepts and terminology and familiarize themselves with rare diagnoses seldom seen in practice, (b) illustrate various types of psychopathology to students and teachers of abnormal psychology as well as other professionals, (c) aid clinicians in studying for

professional examinations, and (d) provide researchers a means of assessing clinicians' levels of diagnostic expertise and reliability. This last recommendation is germane to the purpose of the present study.

The case study utilized by the present study was published by Spitzer et al. in their Psychopathology: A Case Book as an example of Adjustment disorder with depressed mood. No concomitant diagnoses were assigned by Spitzer et al. to this case study. The present study abbreviated this case study by omitting details regarding the chronological progress of therapy. Additionally, information regarding symptomatology was slightly altered to reflect a more serious affective disorder. The goal of this change was to produce a case study that would be reasonably representative of both Adjustment disorder with depressed mood and a more serious diagnosis of Unipolar depression. The resulting case study was presented to subjects with the following instructions:

> This study investigates factors influencing clinical judgment. Your job is to read the following case study and, based on what you have read, estimate the subject's level of functioning and prognosis on the subsequent two rating scales.

Labeling Manipulation

The case study was then manipulated to assess labeling bias and anchoring effects. To assess labeling bias, one of

two diagnostic labels was assigned to the case study subject. Either the sentence "Although his problems may appear minor, he is actually suffering from Unipolar depression," or the sentence "Although his problems may appear major, he is actually struggling with Adjustment disorder with depressed mood," was added to the case study. These two sentences constituted the two levels of the "diagnostic label" factor. This manipulation followed the methodology employed by Temerlin (1968, 1970) to investigate labeling bias effects in clinical judgment.

Anchoring Manipulation

To assess anchoring effects, the timing of this labeling manipulation was varied. The diagnostic label sentences were added either at the beginning of the case study, in the instructions, or at the end of the case study as its final sentence. Varying the location of presentation of the diagnostic label constituted the two levels (early and late) of the "time" factor. This manipulation followed the methodology employed by Friedlander and Stockman (1983), Friedlander and Phillips (1984), and Levin (1984) to investigate anchoring effects in clinical judgment.

In concert with the model constructed by Heverly et al. for effective analogue research stimuli, other than the diagnostic label and time manipulations there were no changes in the altered form of the case study published by Spitzer et al. The four versions of the case study created

by these two manipulations can be seen in Appendices C, D, E, and F.

Pretesting the Manipulations

Validating the case study involved demonstrating the differential impact of the two labels (Adjustment disorder with depressed mood, Unipolar depression) on the dependent variables, and assuring that each label was equally representative of the case study. First, the differential impact of the diagnostic labels was established. Twelve doctoral students from an APA-approved counseling psychology program were asked to rate the level of pathology on the GAS and prognosis on the Prognostic Scale of hypothetical individuals with the following diagnoses: Schizophrenia, paranoid; Somatization disorder; Adjustment disorder with depressed mood; Unipolar depression; Panic disorder; Borderline personality disorder; and Depersonalization disorder. Analysis of variance of mean GAS ratings was significant: $F(6,77) = 11.56$, $p < .0001$. A Newman-Keuls multiple comparison procedure conducted at the .05 level of significance revealed that an individual suffering from Adjustment disorder with depressed mood ($\bar{X} = 64.92$) was perceived as significantly better adjusted than someone suffering from Unipolar depression ($\bar{X} = 32.08$).

Analysis of variance of mean Prognostic Scale ratings was also significant: $F(6,77) = 10.53$, $p < .0001$. A Newman-Keuls multiple comparison procedure conducted at the .05

level of significance revealed that an individual suffering from Adjustment disorder with depressed mood ($\bar{X}$ = 1.50) was perceived as having a significantly better prognosis than someone experiencing Unipolar depression ($\bar{X}$ = 3.08). Thus, the diagnostic labels manipulated by the present study were rated differentially on both the GAS and the Prognostic Scale by pretesting subjects.

The representativeness of the diagnostic labels was then established. Ten doctoral clinicians were asked to read the case study and indicate on a one to seven scale (1 = very poor fit and 7 = very good fit) the goodness-of-fit of the following diagnoses with the case study subject: Schizophrenia, paranoid; Somatization disorder; Adjustment disorder with depressed mood; Unipolar depression; Panic disorder; Borderline personality disorder; Depersonalization disorder; Multiple personality; Conversion disorder; and Bipolar disorder, mixed. Analysis of variance of mean goodness-of-fit ratings for these diagnoses was significant: $F(9,90) = 44.19$, $p < .0001$. A Newman-Keuls multiple comparison procedure conducted at the .05 level of significance revealed that Adjustment disorder with depressed mood ($\bar{X}$ = 5.5) and Unipolar depression ($\bar{X}$ = 4.9) did not differ, but did differ from all other mean goodness-of-fit ratings. Thus, pretesting subjects rated both the diagnostic labels as representative of the case study.

Global Assessment Scale

The Global Assessment Scale (Appendix G) was designed by Endicott, Spitzer, Fleiss, and Cohen (1976) as a rating scale clinicians could use to evaluate an individual's level of pathology on a continuum ranging from mental health to psychological/psychiatric sickness. The GAS is a modification of the Health-Sickness Rating Scale (HSRS), which was constructed by Luborsky (1962) and the Menninger Psychotherapy Project. Thus, an adequate understanding of the GAS requires an understanding of its precursor, the HSRS.

Luborsky and the researchers at the Menninger Psychotherapy Project were interested in investigating whether the concept of "mental health" was tangible enough to be reliably judged along a single continuum by experienced observers. They saw the need for "a simple survey instrument to record shorthand judgments of a patient's status--one that would permit recording of changes over time in a single case and easy comparison of one case with another" (Luborsky, 1962, p. 408). They concluded that assessing "mental health" would best be accomplished by a scale that compared a patient with a standard series of patients previously graded in levels of mental health. Such an instrument, they reasoned, would yield both an initial absolute rating of mental health and, upon retesting, an improvement or deterioration rating of movement across levels of mental health.

The resulting instrument, the HSRS, was a 100-point scale anchored at the lower end (0) with the description Any condition which, if unattended, would quickly result in the patient's death, at the upper end (100) with the description An ideal state of complete functioning integration, resiliency in the face of stress, happiness and social effectiveness, and at the intermediate anchor points 10, 25, 35, 50, 65, and 75 with intermediate descriptions. These seven anchor points created seven levels of mental health/sickness. Descriptions of persons representative of each level were provided on the scale in diagnostic terms, and 34 previously scored case studies were also included as examples to aid in scoring. Thus, rating a patient on the HSRS required a clinician to first decide on the general level of health/sickness (e.g., 50-65) and then assign a specific score (e.g., 58) by comparing the patient to one of the 34 previously scored case studies.

Twelve years following the publication of the HSRS, Luborsky and Bachrach (1974) evaluated 18 studies that had utilized it and concluded that the experiences with its use "bear out its original promise" (Luborsky & Bachrach, 1974, p. 292). According to Luborsky and Bachrach, clinicians had found the HSRS easy to use and valid across the entire range of mental disorders. The interjudge reliability of the HSRS ranged in these 18 studies from .65 to .94, with most of the reliability coefficients distributed toward the upper end of this range, and a test-retest (one year) reliability

coefficient of .77 was also obtained (Luborsky and Bachrach, 1974).

Luborsky and Bachrach also presented correlations between the HSRS and other measures as evidence of concurrent validity: .84 with the Prognostic Index Interview subtotal (Auerbach, Luborsky & Johnson, 1972); .55 with the Jenkins Symptom Rating Scale (Distler, May, & Tuma, 1964); .64 with the same measure (May & Tuma, 1964); and .66 with the Ann Arbor Psychotic Confusion Scale (May & Tuma, 1964). However, correlations of only .20, -.13, and .20 were reported with the Paranoia, Psychasthenia, and Schizophrenia MMPI scales, respectively (May & Tuma, 1964).

Endicott, Spitzer, Fleiss, and Cohen (1976) made four modifications in the HSRS in the development of the GAS. First, all anchor points were reworded in clinical rather than diagnostic terms. Second, the anchor point In need of hospitalization was removed, as Endicott et al. noted that patients with severe conditions were increasingly being treated in the community rather than the hospital. Substituted was the anchor point Any symptomatology that obviously required treatment in order for the person to continue to function adequately. Third, descriptions of specific behaviors representative of each level of health/sickness were substituted for the descriptions of persons representative of each level. Fourth, the 100-point range was divided into ten equal levels (with intermediate anchor points at 10, 20, 30, 40, 50, 60, 70, 80, and 90)

instead of seven unequal levels to provide for greater specification at the upper and lower ends of the scale.

Endicott et al. felt that these modifications were extensive enough to warrant redefining the instrument the Global Assessment Scale, but stressed that the new instrument "retains the basic idea and structure of the HSRS" (Endicott, Spitzer, Fleiss, & Cohen, 1976, p. 766). Endicott et al. conducted five studies assessing the psychometric properties of the GAS prior to publishing it in 1976. Intraclass reliability coefficients in these five studies ranged from .69 to .91, and adequate concurrent validity was reported with the Psychiatric Status Schedule, Family Evaluation Form, and Mental Status Examination Record. Additionally, the GAS was demonstrated by Endicott et al. to have greater sensitivity to change than other global ratings of overall severity of illness. Endicott et al. concluded that "the relative simplicity, reliability, and validity of the GAS suggests that it would be useful in a wide variety of clinical and research settings" (Endicott, Spitzer, Fleiss, & Cohen, 1976, p. 766).

Eight years following its publication, Dekker (1984) examined the psychometric properties of the GAS more closely by evaluating the over 90 studies that had already utilized it. Thirty-one of these studies yielded an acceptable mean interjudge reliability coefficient of .80. As with the HSRS, correlations with the MMPI reflecting concurrent validity were low, ranging from .29 to .36, leading Dekker

to warn against the specific use of the GAS as a cutoff score for individuals. However, Dekker noted that mean GAS scores appeared to have accuracy and concluded that the GAS seemed appropriate for comparing group mean scores. The present study operationally defined group mean GAS scores as a measure of clinical judgment as had Friedlander and Stockman (1983), Friedlander and Phillips (1984), and Levin (1984).

The authors of the GAS presented it with the following instructions:

> Rate the subject's lowest level of functioning in the last week by selecting the lowest range which describes his functioning on a hypothetical continuum of mental health-illness. For example, a subject whose "behavior is considerably influenced by delusions" (range 21-30) should be given a rating in that range even though he has "major impairment in several areas" (range 31-40). Use intermediary levels when appropriate (e.g., 35, 58, 63). Rate actual functioning independent of whether or not subject is receiving and may be helped by medication or some other form of treatment. (Spitzer, Gibbon, & Endicott, 1971, p. 20)

Several changes were made in these instructions by the present study. "John" was substituted for "the subject" to help clinicians and students refer to the case study when making GAS ratings. The phrase "in the last week" and the sentence concerning medication or other treatment were deleted, as this type of information did not appear in the case study selected from Spitzer, Skodol, Gibbon, and Williams' Psychopathology: A Case Book. The 100-point scale was transformed into a 10-point scale, and subjects were instructed to select one figure (e.g., 5) rather than

one range (e.g., 51-60) which best represents John's lowest level of functioning (level of pathology). This last change was deemed necessary after pretesting subjects reported some confusion over whether they should respond with a range rating or a figure rating.

Prognostic Scale

The Prognostic Scale (Appendix H) was derived by Friedlander and Stockman (1983) from Axis V of the DSM-III multiaxial diagnostic system (American Psychiatric Association, 1980). DSM-III presents Axis V as a seven-point scale on which a clinician rates an individual's highest level of adaptive functioning from 1 (Superior) to 7 (Grossly Impaired). A composite of three areas denotes DSM-III's concept of adaptive functioning: social relations, occupational functioning, and use of leisure time. DSM-III reports that two field trials with adult patients resulted in intraclass reliability coefficients for Axis V of .75 and .80 (Spitzer, Forman, & Nee, 1979).

Friedlander and Stockman combined Axis V with the following instructions to comprise the Prognostic Scale:

> Rate the highest level of adaptive functioning that could be expected for this client (i.e., a prognosis), given sufficient motivation for change, a good therapeutic relationship, and adequate time for whatever form of treatment is adopted. (Friedlander & Stockman, 1983, p. 641)

The only change in these instructions by the present study was the substitution of "John" for "this client" to help clinicians and students refer to the case study when

making Prognostic Scale ratings and to make the experimental materials consistent. The present study operationally defined group mean Prognostic Scale scores as a measure of clinical judgment, as had Friedlander and Stockman (1983), Friedlander and Phillips (1984), and Levin (1984).

Following the Prognostic Scale, subjects were asked to indicate the degree of confidence they placed in their GAS and Prognostic Scale ratings on a seven-point scale, with 1 = Not at all confident and 7 = Very confident. The experimental materials then ended by inviting subjects to react to the research (as a manipulation check) and thanking them for their participation. Clinicians were additionally asked to return their completed materials in the self-addressed, stamped envelope provided. Two reminders were mailed to clinicians two and four weeks subsequently to again request participation. These reminders can be seen in Appendices I and J.

Experimental Design

The present study replicated the experimental design Caetano (1974) developed to test labeling theory and added an anchoring manipulation. Caetano theorized that the medical experience of clinicians (psychiatrists) and the suggestion of illness provided by labeling an interview subject "mental patient" would affect diagnosis. Caetano wrote of his research: "The objective of the study was to determine the joint-effects of clinical experience and

suggestion, thereby providing a limited test of the labeling theory assumptions" (Caetano, 1974, p. 254). He concluded that "if these assumptions are valid, we would expect clinical experience and suggestion to increase presumption of mental illness" (Caetano, 1974, p. 254). The purpose of the present study was to add to the clinical judgment literature by investigating whether the judgmental heuristic anchoring qualifies Caetano's study and other studies (Berman & Berman, 1984; Langer & Abelson, 1974; Temerlin, 1968) interpreted as supporting labeling theory.

The present study employed a 2 x 2 x 2 (Experience x Label x Time) factorial design. All factors were between-subjects. The experience factor identified subjects as either experienced (clinicians) or inexperienced (students) with DSM-III diagnostic labels and clinical judgment situations. The label factor referred to the DSM-III diagnoses (Adjustment disorder with depressed mood or Unipolar depression) that were assigned to the case study subject. The time factor referred to the time of presentation of these labels in the case study (early, at the end of the instructions, or late, as the last sentence of the case study). These three factors combined to create eight experimental conditions. Subjects were randomly assigned to one condition by experience level. Subjects contributed data in the form of ratings on two dependent measures of clinical judgment, the Global Assessment Scale and the Prognostic Scale. Subjects also indicated their degree of

confidence in these ratings. Finally, a minimum n of 20 participating subjects in each of the eight experimental conditions was required.

Summary

Studies of clinical judgment often employ analogue research procedures in an attempt to maintain a balance between experimental rigor and clinical reality (Heverly, Fitt, & Newman, 1984). Analogue research uses experimenter-controlled stimuli to simulate clinical encounters and elicits judgments from observers of these encounters. Although analogue research has been sharply criticized regarding generalizability (Goldman, 1978), it has also been strongly supported as a practical means of investigating many clinical situations (Kazdin, 1980). To many researchers, the control of internal validity afforded by analogue research overrides the issue of external validity (Cozby, 1977), which can subsequently be demonstrated through systematic replication. The procedure employed by the present study was modeled after the analogue research procedure Henry and Cohen (1983) utilized to investigate labeling effects in clinical judgment. Henry and Cohen mailed copies of a case study selected from Spitzer, Skodol, Gibbon, and Williams' DSM-III Case Book (1981) to a large number of psychiatrists. The case study subject's sex was alternately labeled as male or female. There were no other

changes in the case study. Labeling bias was successfully identified by Henry and Cohen in the psychiatrists' diagnoses of the case study subject.

In the present study four hundred clinicians, selected by the sampling plan previously outlined, were mailed the experimental materials and invited to participate. Two weeks following this original mailing a brief reminder was mailed to those clinicians whose completed materials had not been received to again request participation. Two more weeks later a second brief reminder, along with a second set of experimental materials, was mailed to those clinicians whose completed materials had still not been received. One month following the date of the mailing of this second reminder the sampling was terminated. This sampling procedure followed commonly accepted strategies for maximizing response rate (Bailey, 1978).

Students enrolled in an undergraduate abnormal psychology class at the University of Florida for the Fall 1986 semester were also invited to participate, and extra credit was offered for participation. Those who agreed to participate were administered the experimental materials in a single group session and given the opportunity to ask questions following completion of the experimental materials. Students with prior clinical experience, including volunteer counseling, were ruled ineligible for participation to preserve the validity of the experience factor.

The experimental materials which clinicians and students received contained a case study selected from Spitzer, Skodol, Gibbon, and Williams' Psychopathology: A Case Book (1983). This case study was manipulated following Heverly, Fitt, and Newman's (1984) model for effective analogue research stimuli. First, the case study subject was assigned one of two diagnostic labels. Second, this diagnostic label manipulation was presented either near the beginning or at the end of the case study. These two manipulations created four versions of the case study. Each subject received, at random, only one version. Subjects were asked to read the case study and estimate the case study subject's level of pathology on the GAS and prognosis on the Prognostic Scale. They were also asked to indicate their degree of confidence in these ratings.

It was hypothesized that subjects' judgments of pathology and prognosis would vary when the case study subject was differentially labeled, and that judgments would also vary according to the subjects' level of clinical experience. It was further hypothesized that the time of presentation of the diagnostic labels would affect judgments, and that this time or "anchoring" effect would interact with both the labeling and experience effects. Finally, it was hypothesized that clinicians would place more confidence in their judgments of pathology and prognosis than students.

CHAPTER IV

RESULTS

Preliminary analyses were conducted to check on the randomization of subjects across experimental conditions, the representativeness of the clinician sample, the relationship between subjects' demographic data and their ratings on the dependent variables, and intercorrelations among the dependent variables. Major analyses were conducted investigating the main and interactive effects of the independent variables on the dependent variables, thereby testing the six hypotheses presented in Chapter I.

All significance tests were two-tailed, with the level of significance set at .05 unless otherwise stated. For analysis-of-variance procedures the method of least squares was the solution for unbalanced data. For *t*-test procedures tests of homogeneity of variance were conducted where necessitated by unbalanced data. Data were analyzed using Statistical Analysis System (SAS) procedures.

Preliminary Analyses

Eighty undergraduate students enrolled in abnormal psychology in the Fall 1986 semester at the University of

Florida participated as subjects in the present study. In addition 400 members of Division 29 (Psychotherapy) of the American Psychological Association were mailed experimental materials and invited to participate. Of the 400 packets of experimental materials mailed to clinicians, 39 were returned to sender as undeliverable. Of the 361 that were not returned as undeliverable, 193 were returned fully completed for a response rate of 53.5%. An additional eight packets were returned partially completed and not used in the data analysis, and the number of packets unsuccessfully delivered to targeted clinicians but not returned to sender is undetermined. The distribution of participating students and clinicians across experimental conditions is depicted in Table 1.

The following demographic data describe the subjects. Students were 70% females (N = 56), 30% males (N = 24). Their modal class standing was Senior (N = 56), modal major was psychology (N = 63), and mean number of prior credits in psychology courses $\bar{X}$ = 22.1. Clinicians were 30.1% females (N = 58), 69.9% males (N = 135), their mean age was $\bar{X}$ = 48.9, and mean years of experience $\bar{X}$ = 16.6.

The randomization of gender across experimental conditions was assessed by a Chi-square statistic comparing the number of male and female students in each experimental condition. The result was nonsignificant, $X^2(3) = 3.33$, $p = .34$, suggesting that randomization of students across experimental conditions was successful. Chi-square

Table 1

Distribution of Subjects in
Experimental Conditions

Experimental Condition	Students	Clinicians
Adjustment disorder with depressed mood X Early	20	53
Adjustment disorder with depressed mood X Late	20	50
Unipolar depression X Early	20	49
Unipolar depression X Late	20	41

comparing the number of respondent and nonrespondent clinicians in each experimental condition was also nonsignificant, $X^2(3) = 4.36$, $p = .23$, suggesting that randomization of clinicians across experimental conditions was successful.

The representativeness of the respondent clinicians to the nonrespondent clinicians was then assessed. Results of t-tests indicated that the mean age ($\bar{X} = 48.9$) of respondents was not significantly different from the mean age ($\bar{X} = 48.3$) of nonrespondents, $t(312) = -.51$, $p = .61$, n.s., and that the mean years of experience ($\bar{X} = 16.6$) of respondents was not significantly different from the mean years of experience ($\bar{X} = 17.2$) of nonrespondents, $t(350) = .53$, $p = .60$, n.s. These results suggest that the 193 clinicians who participated by completing and returning the experimental materials were representative of the larger sample of clinicians from which they were drawn.

Correlations between the subjects' demographic data and their ratings on the dependent variables were computed to assess the possible contribution of subject variables to experimental effects. Intercorrelations among the dependent variables were also computed; it was predicted that correlation of GAS and Prognostic Scale ratings would be significant and negative, replicating Friedlander and Stockman (1983) and Friedlander and Phillips (1984). The Pearson product-moment correlations of demographic data with dependent variables and intercorrelations among the dependent variables are depicted in Table 2. Results indicated that

Table 2

Correlation of Demographic Data with Dependent Variables and Intercorrelation of Dependent Variables

	GAS	Prognostic Scale	Confidence
Clinicians' Age	.18 $p = .02$	.11 $p = .14$	-.06 $p = .44$
Clinicians' Years of Experience	.13 $p = .07$	.12 $p = .09$	-.10 $p = .17$
Students' Class Standing	-.02 $p = .88$	.13 $p = .23$	-.09 $p = .45$
Students' Number of Psychology Credits	-.14 $p = .22$	-.01 $p = .95$	.06 $p = .58$
GAS		-.02 $p = .80$	-.02 $p = .77$
Prognostic Scale			-.24 $p < .001$
Confidence			

the class standing and number of prior psychology credits of the students were not significantly correlated with their ratings on the GAS, Prognostic Scale, or level of confidence ($p > .01$), and that the age and years of experience of the clinicians were not significantly correlated with their ratings on the GAS, Prognostic Scale, or level of confidence ($p > .01$). Intercorrelations among the dependent variables revealed a significant negative correlation between Prognostic Scale and level of confidence ratings ($r = -.24$, $p < .001$). All other intercorrelations were nonsignificant ($p > .01$).

In addition to the analyses summarized in Table 2, t-tests were computed to assess the relationship between subjects' sex and their ratings on the dependent variables. Results indicated that male students' mean GAS ratings ($\bar{X} = 5.88$) did not differ significantly from female students' mean GAS ratings ($\bar{X} = 5.96$), $t(78) = .34$, $p = .74$, n.s., and that male students' mean Prognostic Scale ratings ($\bar{X} = 2.13$) did not differ significantly from female students' mean Prognostic Scale ratings ($\bar{X} = 2.07$), $t(78) = -.23$, $p = .81$, n.s. Similarly, male clinicians' mean GAS ratings ($\bar{X} = 5.89$) did not differ significantly from female clinicians' mean GAS ratings ($\bar{X} = 5.84$), $t(191) = -.27$, $p = .79$, n.s., and male clinicians' mean Prognostic Scale ratings ($\bar{X} = 2.13$) did not differ significantly from female clinicians' mean Prognostic Scale ratings ($\bar{X} = 2.16$), $t(191) = .25$, $p = .81$, n.s.

Major Analyses

Subjects' ratings on the three dependent variables were analyzed as follows. A 2 x 2 x 2 (Label x Experience x Time) analysis-of-variance procedure was computed to investigate the main and interactive effects of the independent variables on GAS ratings. A similar 2 x 2 x 2 analysis-of-variance procedure was computed to investigate the main and interactive effects of the independent variables on Prognostic Scale ratings. A *t*-test procedure was computed comparing students' and clinicians' mean ratings of the level of confidence they placed in their GAS and Prognostic Scale ratings.

Global Assessment Scale

The 2 x 2 x 2 ANOVA along the GAS revealed a significant main effect for label, $F(1,265) = 6.55$, $p < .05$, though no other effects were found, as depicted in Table 3. The means and standard deviations of GAS ratings are depicted in Table 4. The significant label main effect revealed that mean GAS ratings for the case study subject labeled with Adjustment disorder with depressed mood ($\bar{X} = 6.10$) were significantly higher than mean GAS ratings for the case study subject labeled with Unipolar depression ($\bar{X} = 5.67$).

Table 3

Analysis of Variance for GAS Ratings

Source of Variation	SS	df	MS	F	p
Label (A)	6.95	1	6.95	6.55	.011
Experience (B)	.34	1	.34	.32	.57
Time (C)	.007	1	.007	.01	.93
A x B	1.75	1	1.75	1.65	.20
A x C	.0003	1	.0003	.00	.99
B x C	.01	1	.01	.01	.92
A x B x C	1.68	1	1.68	1.58	.21
Error	281.21	265	1.06		

Table 4

Means and Standard Deviations of GAS Ratings

	Experience			
	Student		Clinician	
Label	Early	Late	Early	Late
Adjustment disorder with depressed mood	5.95 (1.00)	6.10 (1.07)	6.21 (.79)	6.04 (.97)
Unipolar depression	5.95 (1.15)	5.75 (1.12)	5.51 (1.19)	5.68 (1.06)

Note: Higher ratings indicate more favorable level of functioning.

Prognostic Scale

The 2 x 2 x 2 ANOVA along Prognostic Scale scores revealed no significant effects, as depicted in Table 5. The means and standard deviations of Prognostic Scale ratings are depicted in Table 6.

Confidence Rating

The *t*-test procedure comparing students' and clinicians' mean ratings of the level of confidence they placed in their GAS and Prognostic Scale ratings was nonsignificant, $t(271) = .32$, $p = .75$, n.s. Students' mean confidence rating was $\bar{X} = 5.38$, SD = 1.01. Clinicians' mean confidence rating was $\bar{X} = 5.42$, SD = 1.07.

Summary of Tests of Hypotheses

1. Assigning varying diagnostic labels (Adjustment disorder with depressed mood or Unipolar depression) to a case study subject will affect evaluations of the subject's level of pathology on the GAS and prognosis on the Prognostic Scale. Partially supported.

2. The level of clinical experience of the evaluators (clinician or student) will affect their evaluations of the case study subject's level of pathology on the GAS and prognosis on the Prognostic Scale. Not supported.

3. Varying the time of presentation of the diagnostic labels (early or late) in the case study will affect

Table 5

Analysis of Variance for
Prognostic Scale Ratings

Source of Variation	SS	df	MS	F	p
Label (A)	.09	1	.09	.14	.71
Experience (B)	.14	1	.14	.22	.64
Time (C)	.67	1	.67	1.01	.32
A x B	.24	1	.24	.37	.54
A x C	.002	1	.002	.00	.96
B x C	1.55	1	1.55	2.34	.13
A x B x C	.80	1	.80	1.20	.27
Error	175.74	265	.66		

Table 6

Means and Standard Deviations of
Prognostic Scale Ratings

	Experience			
	Student		Clinician	
Label	Early	Late	Early	Late
Adjustment disorder with depressed mood	1.90 (.72)	2.30 (1.08)	2.17 (.87)	2.00 (.70)
Unipolar depression	2.00 (1.03)	2.15 (.88)	2.16 (.66)	2.22 (.79)

Note: Lower ratings indicate more favorable prognosis.

evaluations of the subject's level of pathology on the GAS and prognosis on the Prognostic Scale. Not supported.

4. The diagnostic labels assigned to the case study subject will interact with the time of presentation of the labels to affect evaluations of the subject's level of pathology on the GAS and prognosis on the Prognostic Scale. Not supported.

5. The level of clinical experience of the evaluators will interact with the time of presentation of the diagnostic labels in the case study to affect evaluations of the subject's level of pathology on the GAS and prognosis on the Prognostic Scale. Not supported.

6. Clinicians will place more confidence than students in their ratings of level of pathology on the GAS and prognosis on the Prognostic Scale. Not supported.

CHAPTER V

DISCUSSION

This chapter presents a discussion of the results. The first section considers theoretical and practical implications of the results. The second section identifies limitations of the study. The third section outlines areas of continued research suggested by the study.

Implications

Hypothesis one was partially supported, as the main effect for label was significant on the GAS but nonsignificant on the Prognostic Scale. This result replicates previous studies (Berman & Berman, 1984; Caetano, 1974; Langer & Abelson, 1974; Temerlin, 1968) identifying labels as potential sources of bias in clinical judgment and extends them to more specific labels. Whereas previous studies manipulated only global labels such as "psychotic" and "normal" or "job applicant" and "patient," the present study demonstrated that specific DSM-III diagnoses are also potential sources of bias in clinical judgment, at least in judgments concerning level of pathology. Practically, these results suggest that clinicians need to recognize and account for the impact of preexisting diagnostic labels when

making judgments of pathology in the clients they evaluate or treat.

Hypothesis two was not supported on either the GAS or Prognostic Scale. The level of clinical experience of the subjects did not significantly affect their ratings on either the GAS or Prognostic Scale. Thus, although the label main effect on the GAS noted above is consistent with labeling theory, labeling theory's assertion that both labels and clinical experience are sources of bias in clinical judgment (Caetano, 1974) was not supported.

Hypotheses three, four, and five were not supported on either the GAS or Prognostic Scale. Varying the time of presentation of diagnostic labels--the anchoring manipulation--did not significantly affect ratings on either the GAS or Prognostic Scale. Nor did time of presentation interact significantly with the diagnostic label manipulation or the level of clinical experience of the subjects. Thus, the prediction that anchoring contributes to labeling bias and clinical experience effects in clinical judgment was not supported. However, these results do not indicate that clinicians are not susceptible to biases due to utilization of judgmental heuristics as outlined by Tversky and Kahneman (1974). Since clear biases did not occur, their causes cannot be identified. The hypothesis that heuristics contribute to labeling bias and clinical experience effects, thereby qualifying labeling theory, must still be considered tenable.

In fact, the largely nonsignificant results of the present study suggest that the representativeness heuristic (Tversky & Kahneman, 1974) may have been the mechanism in operation in the previous studies that reported labeling bias and clinical experience effects. Representativeness is the perceived degree to which A resembles B (Tversky & Kahneman, 1974). Failure to control the representativeness of manipulated labels to stimulus (case study, videotape, or audiotape) subjects may have contributed to the significant labeling bias and clinical experience effects found by Berman and Berman (1984), Caetano (1974), Langer and Abelson (1974), and Temerlin (1968). For example, Berman and Berman's (1984) manipulation of the label "psychotic" was inconsistent with the stimulus subject--a normal individual--this label was assigned to. Psychotic behavior was not indicated in this individual. Similar inconsistencies between labeling manipulations and stimulus subjects may have contributed to the other labeling bias and clinical experience effects reported.

Conversely, the control of representativeness afforded via pretesting in the present study may explain why these effects were not replicated. It seemed logical to conclude that global, vastly disparate labels such as "psychotic" and "normal" could scarcely be equally representative of any stimulus subject. Therefore, the present study developed a case study which was demonstrated to equally represent the two labels chosen for manipulation. The resulting

consistency between the labeling manipulations and case study information may have "washed out" labeling bias and clinical experience effects. Practically, the need for analogue researchers to ensure the representativeness of the labels they manipulate to stimulus subjects is highlighted by these results.

Hypothesis six was not supported. There was no significant difference in the amount of confidence clinicians and students placed in their GAS and Prognostic Scale ratings. Previous findings that clinicians place more confidence in their judgments than students (Friedlander & Phillips, 1984) and that extreme confidence accompanies erroneous judgments (Fischoff, Slovic, & Lichtenstein, 1977) were not replicated.

The significant negative correlation between GAS and Prognostic Scale ratings found by Friedlander and Stockman (1983) and Friedlander and Phillips (1984) was not replicated by the present study. This result may bring into question the use of these two dependent measures as definitive measures of clinical judgment, since logically as pathology increases prognosis should always decrease (i.e., become less optimistic), and vice versa. An unexpected finding was a significant negative correlation ($r = -.24$) between Prognostic Scale ratings and level of confidence ratings. This result indicates that subjects were more confident of poorer prognoses and less confident of more optimistic prognoses.

Limitations

The following limitations accompany the present study. First, as was the case with previous research, knowing that their judgments were being evaluated may have led subjects to respond in what they perceived to be a socially desirable fashion rather than report their true impressions. Second, clinicians may have had some familiarity with the GAS and/or Prognostic Scale and a predilection to respond to either or both in a biased manner. (Although this possibility is not supported by the lack of significant differences between clinicians' ratings and those of students unfamiliar with the GAS and Prognostic Scale.) Third, the contribution of different data collection methods (clinicians by mail and students in person) to experimental effects is undetermined. (Although, again, there were no significant differences between clinicians' and students' ratings.) Fourth, the assumption that ordinal judgments on the GAS and Prognostic Scale accurately reflect actual clinical practice--where categorical judgments such as diagnoses are often made--may not be valid.

Fifth, regarding generalizability to other persons, these results are limited to the extent that 1) Division 29 members represent all clinicians, and 2) abnormal psychology students represent all nonexpert judges. Sixth, generalizability to real-life counseling situations can only be inferred due to the analogue research methodology employed. Although generalizability is suggested by similar previous

research, it is nonetheless important to repeat that any artificial diagnostic situation such as that manipulated by the present study is valid only insofar as it accurately mimics actual clinical encounters.

Suggestions for Further Research

The inability of the present study to replicate the anchoring effect found by Friedlander and Stockman (1983) made it impossible to assess the relationship between anchoring and other independent variables. Apparently Friedlander and Stockman's longer, more florid labeling manipulation elicited a more potent, robust effect than did the brief, more benign manipulation utilized in the present study. Future research should recognize that to obtain anchoring effects additional pathognomonic information should be added to the one-sentence manipulation utilized here. More research is needed to establish the amount and type of information necessary to obtain anchoring effects in clinical judgment.

The present study suggested that previous failures to control the representativeness of manipulated labels to stimulus subjects may have contributed to findings of labeling bias and clinical experience effects. This failure to control representativeness occurs in a great deal of analogue research. Future analogue research should attempt, through pretesting or more careful development of independent variables, to demonstrate that the levels of each

planned experimental manipulation equally represent each respective stimulus subject. Such control of representativeness would allow actual effects to be uncovered that previously have been masked or, as was the case here, suggest that representativeness contributed to effects previously thought robust.

The validity of the GAS and Prognostic Scale as measures of clinical judgment and the relationship between these two measures is worthy of additional research. Formal theories and models of the clinical judgment process are constantly evolving (Wiggins, 1981) which may already be identifying dependent variables superior to the GAS or Prognostic Scale. Or, perhaps as diagnostic systems continue to be refined and improved beyond DSM-III no dependent measures of clinical judgment will be considered adequate save diagnoses. In any event, our understanding of the complex cognitive process we call clinical judgment could still be called crude, and research attempting to more adequately define and assess clinical judgment remains necessary.

Finally, the potential relationship between judgmental heuristics and the labeling theory of mental illness remains undetermined. Replication and extension of the present study across a variety of analogue research stimuli and with creative manipulations of anchoring, representativeness, and availability is needed. Such research should speak to the still ongoing debate regarding labeling theory and help

reveal the effects of clinicians' utilization of judgmental heuristics. It remains a persistent irony that most clinicians spend a great deal of time evaluating the cognitive processes of others and so little time considering what systematic errors and biases they themselves are subject to (cf. Friedlander & Stockman, 1983).

APPENDIX A

COVER SHEET FOR CLINICIANS

UNIVERSITY OF FLORIDA

Gainesville, 32611

I am conducting as my doctoral dissertation a study investigating factors influencing clinical judgment. As an experienced clinician you have been carefully selected to participate. Your cooperation is important to the successful completion of the project. Realizing that your time is very valuable, the materials have been organized to take approximately 10 minutes to complete.

There are no known risks associated with participation, no monetary compensation is involved, and your participation will be kept completely confidential by a numbering system in the event results are published. You are free to withdraw your participation at any time, without prejudice.

Please feel free to contact me if you have any questions. Again, your participation is greatly appreciated.

Sincerely,

Curtis Walling
Principal Investigator
University of Florida
Box 110, Psychology
Gainesville, FL 32611

DEMOGRAPHIC INFORMATION

Your age:_____ Sex:_____Male _____Female Years of experience:_____

APPENDIX B

COVER SHEET FOR STUDENTS

UNIVERSITY OF FLORIDA

Gainesville, 32611

I am conducting as my doctoral dissertation a study investigating factors influencing clinical judgment. Your job is to complete the judgment tasks which follow this cover sheet. Your cooperation is important to the successful completion of the project.

There are no known risks associated with participation, no monetary compensation is involved, and your participation will be kept completely confidential (do not write your name on the materials). You are free to withdraw your participation at any time, without prejudice.

Please feel free to ask any questions you might have. Your participation is greatly appreciated.

Sincerely,

Curtis Walling
Principal Investigator
University of Florida
Box 110, Psychology
Gainesville, FL 32611

DEMOGRAPHIC INFORMATION

Your Class_____Fr.
_____Soph.
_____Jr.
_____Sr.

Your Major: _____

Number of prior psychology credits: _____

Sex:_____Male
_____Female

APPENDIX C

CASE STUDY, ADJUSTMENT DISORDER WITH DEPRESSED MOOD X EARLY CONDITION

INSTRUCTIONS

This study investigates factors influencing clinical judgment. Your job is to read the following case study and, based on what you have read, estimate the subject's level of functioning and prognosis on the subsequent two rating scales. Although his problems may appear major, he is actually struggling with Adjustment disorder with depressed mood.

CASE STUDY

John was a 21-year-old senior in a southwestern University. He requested treatment at the university counseling center because of difficulties that he had been having since his girlfriend, Jill, broke up with him three months previously. The relationship with Jill had been "a real love relationship." Nonetheless, they decided to date other people, and eventually she came to prefer another man.

Since the breakup John had lost interest in his schoolwork, was not able to concentrate on his studies, and as a result had accumulated several "incompletes." He also had trouble sleeping and was preoccupied with thoughts about Jill, for whom he still cared. When thinking about her he felt sad and was no longer able to enjoy himself with friends or to pursue his interests in sports and politics. He now thinks he is a failure at school and in his relationships and wonders if he can go on. He feels immobilized and "down in the dumps" because no one really cares about him or understands how he feels about his girlfriend.

Born in a southern state, the second son of a working-class family, John described a family relationship that was harmonious but not close. His father was hardworking and "loyal to his family." The only argument he remembered ever having with his father concerned his decision to grow a beard--to which his father strongly objected. He described his mother as "outgoing" but added "for a few years there she was very dissatisfied and depressed about life." He had positive feelings toward both parents, whom he felt were basically loving and supportive. John's 32-year-old brother was a highly successful lawyer to whom John was not particularly close.

During his high school and college years John was active in athletics and student affairs; he enjoyed other people's company and was well liked by his peers. He had worked part-time as a busboy in the campus coffee shop to help support himself and was getting good grades in all his courses until his current difficulties. He had dated several other girls before his relationship with Jill, but had "never been in love before."

APPENDIX D

CASE STUDY, ADJUSTMENT DISORDER WITH DEPRESSED MOOD X LATE CONDITION

INSTRUCTIONS

This study investigates factors influencing clinical judgment. Your job is to read the following case study and, based on what you have read, estimate the subject's level of functioning and prognosis on the subsequent two rating scales.

CASE STUDY

John was a 21-year-old senior in a southwestern University. He requested treatment at the university counseling center because of difficulties that he had been having since his girlfriend, Jill, broke up with him three months previously. The relationship with Jill had been "a real love relationship." Nonetheless, they decided to date other people, and eventually she came to prefer another man.

Since the breakup John had lost interest in his schoolwork, was not able to concentrate on his studies, and as a result had accumulated several "incompletes." He also had trouble sleeping and was preoccupied with thoughts about Jill, for whom he still cared. When thinking about her he felt sad and was no longer able to enjoy himself with friends or to pursue his interests in sports and politics. He now thinks he is a failure at school and in his relationships and wonders if he can go on. He feels immobilized and "down in the dumps" because no one really cares about him or understands how he feels about his girlfriend.

Born in a southern state, the second son of a working-class family, John described a family relationship that was harmonious but not close. His father was hardworking and "loyal to his family." The only argument he remembered ever having with his father concerned his decision to grow a beard--to which his father strongly objected. He described his mother as "outgoing" but added "for a few years there she was very dissatisfied and depressed about life." He had positive feelings toward both parents, whom he felt were basically loving and supportive. John's 32-year-old brother was a highly successful lawyer to whom John was not particularly close.

During his high school and college years John was active in athletics and student affairs; he enjoyed other people's company and was well liked by his peers. He had worked part-time as a busboy in the campus coffee shop to help support himself and was getting good grades in all his courses until his current difficulties. He had dated several other girls before his relationship with Jill, but had "never been in love before." Although his problems may appear major, he is actually struggling with Adjustment disorder with depressed mood.

APPENDIX E

CASE STUDY, UNIPOLAR DEPRESSION X EARLY CONDITION

INSTRUCTIONS

This study investigates factors influencing clinical judgment. Your job is to read the following case study and, based on what you have read, estimate the subject's level of functioning and prognosis on the subsequent two rating scales. Although his problems may appear minor, he is actually suffering from Unipolar depression.

CASE STUDY

John was a 21-year-old senior in a southwestern University. He requested treatment at the university counseling center because of difficulties that he had been having since his girlfriend, Jill, broke up with him three months previously. The relationship with Jill had been "a real love relationship." Nonetheless, they decided to date other people, and eventually she came to prefer another man.

Since the breakup John had lost interest in his schoolwork, was not able to concentrate on his studies, and as a result had accumulated several "incompletes." He also had trouble sleeping and was preoccupied with thoughts about Jill, for whom he still cared. When thinking about her he felt sad and was no longer able to enjoy himself with friends or to pursue his interests in sports and politics. He now thinks he is a failure at school and in his relationships and wonders if he can go on. He feels immobilized and "down in the dumps" because no one really cares about him or understands how he feels about his girlfriend.

Born in a southern state, the second son of a working-class family, John described a family relationship that was harmonious but not close. His father was hardworking and "loyal to his family." The only argument he remembered ever having with his father concerned his decision to grow a beard--to which his father strongly objected. He described his mother as "outgoing" but added "for a few years there she was very dissatisfied and depressed about life." He had positive feelings toward both parents, whom he felt were basically loving and supportive. John's 32-year-old brother was a highly successful lawyer to whom John was not particularly close.

During his high school and college years John was active in athletics and student affairs; he enjoyed other people's company and was well liked by his peers. He had worked part-time as a busboy in the campus coffee shop to help support himself and was getting good grades in all his courses until his current difficulties. He had dated several other girls before his relationship with Jill, but had "never been in love before."

APPENDIX F

CASE STUDY, UNIPOLAR DEPRESSION X LATE CONDITION

INSTRUCTIONS

This study investigates factors influencing clinical judgment. Your job is to read the following case study and, based on what you have read, estimate the subject's level of functioning and prognosis on the subsequent two rating scales.

CASE STUDY

John was a 21-year-old senior in a southwestern University. He requested treatment at the university counseling center because of difficulties that he had been having since his girlfriend, Jill, broke up with him three months previously. The relationship with Jill had been "a real love relationship." Nonetheless, they decided to date other people, and eventually she came to prefer another man.

Since the breakup John had lost interest in his schoolwork, was not able to concentrate on his studies, and as a result had accumulated several "incompletes." He also had trouble sleeping and was preoccupied with thoughts about Jill, for whom he still cared. When thinking about her he felt sad and was no longer able to enjoy himself with friends or to pursue his interests in sports and politics. He now thinks he is a failure at school and in his relationships and wonders if he can go on. He feels immobilized and "down in the dumps" because no one really cares about him or understands how he feels about his girlfriend.

Born in a southern state, the second son of a working-class family, John described a family relationship that was harmonious but not close. His father was hardworking and "loyal to his family." The only argument he remembered ever having with his father concerned his decision to grow a beard--to which his father strongly objected. He described his mother as "outgoing" but added "for a few years there she was very dissatisfied and depressed about life." He had positive feelings toward both parents, whom he felt were basically loving and supportive. John's 32-year-old brother was a highly successful lawyer to whom John was not particularly close.

During his high school and college years John was active in athletics and student affairs; he enjoyed other people's company and was well liked by his peers. He had worked part-time as a busboy in the campus coffee shop to help support himself and was getting good grades in all his courses until his current difficulties. He had dated several other girls before his relationship with Jill, but had "never been in love before." Although his problems may appear minor, he is actually suffering from Unipolar depression.

APPENDIX G

GLOBAL ASSESSMENT SCALE

Rate John's lowest level of functioning by selecting the lowest figure which describes his functioning on a hypothetical continuum of mental health-illness. For example, an individual whose "behavior is considerably influenced by delusions" should be given a (3) rating even though he has "major impairment in several areas" (a 4 rating).
(circle one)

10 – No symptoms, superior functioning in a wide range of activities, life's problems never seem to get out of hand, is sought out by others because of his warmth and integrity.

9 – Transient symptoms may occur, but good functioning in all areas, interested and involved in a wide range of activities, socially effective, generally satisfied with life, "everyday" worries that only occasionally get out of hand.

8 – Minimal symptoms may be present but no more than slight impairment in functioning, varying degrees of "everyday" worries and problems that sometimes get out of hand.

7 – Some mild symptoms (eg, depressive mood and mild insomnia) OR some difficulty in several areas of functioning, but generally functioning pretty well, has some meaningful interpersonal relationships and most untrained people would not consider him "sick."

6 – Moderate symptoms OR generally functioning with some difficulty (eg, few friends and flat affect, depressed mood, and pathological self-doubt, euphoric mood and pressure of speech, moderately severe antisocial behavior).

5 – Any serious symptomatology or impairment in functioning that most clinicians would think obviously requires treatment or attention (eg, suicidal preoccupation or gesture, severe obsessional rituals, frequent anxiety attacks, serious antisocial behavior, compulsive drinking).

4 – Major impairment in several areas, such as work, family relations, judgment, thinking, or mood (eg, depressed woman avoids friends, neglects family, unable to do housework), OR some impairment in reality testing or communication (eg, speech is at times obscure, illogical, or irrelevent), OR single serious suicide attempt.

3 – Unable to function in almost all areas (eg, stays in bed all day), OR behavior is considerably influenced by either delusions or hallucinations, OR serious impairment in communication (eg, sometimes incoherent or unresponsive) or judgment (eg, acts grossly inappropriately).

2 – Needs some supervision to prevent hurting self or others, or to maintain minimal personal hygiene (eg, repeated suicide attempts, frequently violent, manic excitement, smears feces), OR gross impairment in communication (eg, largely incoherent or mute).

1 – Needs constant supervision for several days to prevent hurting self or others, or makes no attempt to maintain minimal personal hygiene.

APPENDIX H

PROGNOSTIC SCALE

Rate the highest level of adaptive functioning that could be expected for John (i.e., a prognosis), given sufficient motivation for change, a good therapeutic relationship, and adequate time for whatever treatment is adopted. (circle one)

1 SUPERIOR - Unusually effective functioning in social relations, occupational functioning, and use of leisure time

2 VERY GOOD - Better than average functioning in social relations, occupational functioning, and use of leisure time

3 GOOD - No more than slight impairment in either social or occupational functioning

4 FAIR - Moderate impairment in either social relations or occupational functioning, or some impairment in both

5 POOR - Marked impairment in either social relations or occupational functioning, or moderate impairment in both

6 VERY POOR - Marked impairment in both social relations and occupational functioning

7 GROSSLY IMPAIRED - Gross impairment in virtually all areas of functioning

---

Please indicate your level of confidence in these ratings of level of functioning and prognosis: (circle one)

1	2	3	4	5	6	7
Not at all confident						Very confident

THANK YOU FOR YOUR PARTICIPATION. FEEL FREE TO USE THE REVERSE SIDE OF THIS PAGE TO RESPOND TO THE RESEARCH. PLEASE RETURN THE MATERIALS IN THE ENVELOPE PROVIDED.

APPENDIX I

FIRST REMINDER TO CLINICIANS

UNIVERSITY OF FLORIDA

Gainesville, 32611

Dear Clinician:

Approximately two weeks ago you received a request to participate in a study investigating factors influencing clinical judgment. As of yet your materials have not been received. Though your time is no doubt very limited, participation will only require 10 minutes.

Your contribution as an experienced clinician is invaluable. Please disregard this reminder should it have crossed your materials in the mail.

Your cooperation is greatly appreciated.

Sincerely,

Curtis Walling
Principal Investigator
University of Florida
Box 110, Psychology
Gainesville, FL 32611

APPENDIX J

SECOND REMINDER TO CLINICIANS

UNIVERSITY OF FLORIDA

Gainesville, 32611

Dear Clinician:

Recently you received a request to participate in a study investigating factors that influence clinical judgment. As of yet your materials have not been received. There is still time for your materials to be included in the data analysis.

As an experienced clinician, your contribution to my doctoral dissertation is invaluable. If you could take the time (10 minutes) to complete and return the materials it would be greatly appreciated.

Thank you very much.

Sincerely,

Curtis Walling
Principal Investigator
University of Florida
Box 110, Psychology
Gainesville, FL 32611

BIBLIOGRAPHY

Ajzen, I. (1977). Intuitive theories of events and the effects of base rate information on prediction. Journal of Personality and Social Psychology, 35, 303-314.

Ajzen, I., Dalto, C.A., & Blyth, D.P. (1979). Consistency and bias in the attribution of attitudes. Journal of Personality and Social Psychology, 37, 1871-1876.

American Psychiatric Association. (1952). Diagnostic and statistical manual of mental disorders. Washington, DC: Author.

American Psychiatric Association. (1968). Diagnostic and statistical manual of mental disorders (2nd ed.). Washington, DC: Author.

American Psychiatric Association. (1980). Diagnostic and statistical manual of mental disorders (3rd ed.). Washington, DC: Author.

Anderson, N.H. (1965). Primacy effects in personality impression formation using a generalized order effect paradigm. Journal of Personality and Social Psychology, 2, 1-9.

Anderson, N.H. (1970). Functional measurement and psychophysical judgment. Psychological Review, 77, 153-170.

Arkes, H.R. (1981). Impediments to accurate clinical judgment and possible ways to minimize their impact. Journal of Consulting and Clinical Psychology, 49, 323-330.

Arkes, H.R., & Harkness, A.R. (1980). The effects of making a diagnosis on subsequent recognition of symptoms. Journal of Experimental Psychology: Human Learning and Memory, 6, 568-575.

Arkes, H.R., Wortmann, R.L., Saville, R.D., & Harkness, A.R. (1981). The hindsight bias among physicians weighing the likelihood of a diagnosis. Journal of Applied Psychology, 66, 252-254.

Asch, S.E. (1946). Forming impressions of personality. Journal of Abnormal and Social Psychology, 41, 258-290.

Auerbach, A.H., Luborsky, L., & Johnson, M. (1972). Clinicians' predictions of psychotherapy outcome: A trial of a prognostic index. American Journal of Psychiatry, 128, 830-835.

Bailey, K.D. (1978). Methods of social research. London: Free Press.

Becker, H.S. (1963). Outsiders: Studies in the sociology of deviance. New York: Free Press.

Beckman, L. (1970). Effects of students' performance on teachers' and observers' attributions of causality. Journal of Educational Psychology, 61, 75-82.

Begelman, D.A. (1975). Ethical and legal issues in behavior modification. In M. Hersen, R.M. Eisler, & P.M. Miller (Eds.), Progress in behavior modification (Vol. 1). New York: Academic Press.

Berger, P. (1978). Medical treatment of mental illness: Psychopharmacological therapeutics revolutionize psychiatric care and present scientific and ethical challenges to society. Science, 200, 974-981.

Bergin, A.E., & Lambert, M.J. (1980). The evaluation of therapeutic outcomes. In A.E. Bergin & S.L. Garfield (Eds.), Handbook of psychotherapy and behavior change. New York: Wiley.

Berman, G.J., & Berman, D.S. (1984). In the eyes of the beholder: Effects of psychiatric labels and training on clinical judgments. Academic Psychology Bulletin, 6, 37-42.

Bieri, J., Atkins, A.L., Briar, S., Leaman, R.L., Miller, H., & Tripodi, T. (1966). Clinical and social judgment: The discrimination of behavioral information. New York: Wiley.

Bieri, J., Orcutt, B.A., & Leaman, R. (1963). Anchoring effects in sequential clinical judgments. Journal of Abnormal and Social Psychology, 67, 616-623.

Braginsky, B., Braginsky, D., & Ring, K.L. (1969). Methods of madness: The mental hospital as a last resort. New York: Holt, Rinehart, & Winston.

Bruner, J.S., & Tagiuri, R. (1954). The perception of people. In G. Lindzey (Ed.), Handbook of social psychology (Vol. 2). Cambridge, MA: Addison-Wesley.

Caetano, D.F. (1974). Labeling theory and the presumption of mental illness in diagnosis: An experimental design. Journal of Health and Social Behavior, 14, 253-260.

Calicchia, J.P. (1981). Attitudinal comparison of mental health professionals toward ex-mental patients. Journal of Psychology, 108, 35-41.

Campbell, D.T., Hunt, W.A., & Lewis, N.A. (1957). The effects of assimilation and contrast in judgments of clinical materials. American Journal of Psychology, 70, 347-360.

Castellan, N.J. (1977). Decision making with multiple probabilistic cues. In N.J. Castellan, D.B. Pisoni, & G.R. Potts (Eds.), Cognitive theory (Vol. 2). Hillsdale, NJ: Erlbaum.

Chapman, L., & Chapman J. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. Journal of Abnormal Psychology, 74, 271-280.

Cline, T. (1985). Clinical judgment in context: A review of situational factors in person perception during clinical interviews. Journal of Child Psychology and Psychiatry, 26, 369-380.

Conover, P. (1976). A reassessment of labeling theory: A constructive response to criticism. In L. Coser & O. Larsen (Eds.), The uses of controversy in sociology. New York: Free Press.

Cozby, P.C. (1977). Methods in behavioral research. Palo Alto, CA: Mayfield.

Crane, L., Zonana, H., & Wiser, S. (1977). Implications of the Donaldson decision: A model for periodic review of committed patients. Hospital and Community Psychiatry, 28, 827-833.

Crumpton, E., Weinstein, A.D., Acker, C.W., & Annis, A.P. (1967). How patients and normals see the mental patient. Journal of Clinical Psychology, 23, 46-49.

Dailey, C.A. (1952). The effects of premature conclusions upon the acquisition of understanding about a person. Journal of Psychology, 33, 133-152.

Davis, N.J. (1972). Labelling theory in deviance research: A critique and reconsideration. Sociological Quarterly, 13, 447-474.

Dawes, R.M. (1976). Shallow psychology. In J.S. Carroll & J.W. Payne (Eds.), Cognitive and social behavior. Hillsdale, NJ: Erlbaum.

Deinhardt, C.L. (1983). Personality assessment and psychological interpretation. Springfield, IL: Thomas.

Dekker, D.J. (1984). A study of the validity of the Global Assessment Scale (Doctoral dissertation, Western Michigan University, 1983). Dissertation Abstracts International, 45, 1010B.

De Rivera, J.H. (1968). The psychological dimension of foreign policy. Columbus, OH: Merrill.

Distler, L.S., May, P.R., & Tuma, A.H. (1964). Anxiety and ego strength as predictors of response to treatment in schizophrenic patients. Journal of Consulting Psychology, 28, 170-177.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), Formal representation of human judgment. New York: Wiley.

Einhorn, H.J. (1972). Expert measurement and mechanical combination. Organizational Behavior and Human Performance, 7, 86-106.

Einhorn, H.J., & Hogarth, R.M. (1978). Confidence in judgment: Persistence of the illusion of validity. Psychological Review, 85, 395-416.

Eisler, R.M., & Polak, P.R. (1971). Social stress and psychiatric disorder. Journal of Nervous and Mental Disease, 153, 227-233.

Endicott, J., Spitzer, R.L., Fleiss, J.L., & Cohen, J. (1976). The Global Assessment Scale: A procedure for measuring overall severity of psychiatric disturbance. Archives of General Psychiatry, 33, 766-771.

Erikson, K.T. (1962). Notes on the sociology of deviance. Social Problems, 9, 307-314.

Farina, A. (1981). Sex and the stigma of mental disorders. Clinical Psychology Review, 1, 223-243.

Feigelson, E., Davis, E., Mackinnon, R., Shands, H., & Schwartz, C. (1978). The decision to hospitalize. American Journal of Psychiatry, 35, 354-357.

Fischoff, B. (1975). Hindsight does not equal foresight: The effect of outcome knowledge on judgment under uncertainty. Journal of Experimental Psychology: Human Perception and Performance, 1, 288-299.

Fischoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. Journal of Experimental Psychology: Human Perception and Performance, 3, 552-564.

Fiske, D.W. (1976). Strategies for personality research. San Francisco: Jossey-Bass.

Franchia, J., Canale, D., Cambria, E., Ruest, E., & Sheppard, C. (1976). Public views of ex-mental patients: A note on perceived dangerousness and unpredictability. Psychological Reports, 38, 495-498.

Franz, S.I. (1920). Handbook of mental examination methods. New York: Macmillan.

Friedlander, M.L., & Phillips, S.D. (1984). Preventing anchoring errors in clinical judgment. Journal of Consulting and Clinical Psychology, 52, 366-371.

Friedlander, M.L., & Stockman, S.J. (1983). Anchoring and publicity effects in clinical judgment. Journal of Clinical Psychology, 39, 637-643.

Galen, R.S., & Gambino, S.R. (1975). Beyond normality: The predictive values and efficiency of medical diagnoses. New York: Wiley.

Garfinkel, H. (1956). Conditions of successful degradation ceremonies. American Journal of Sociology, 61, 420-424.

Gershon, E., Bunney, W., & Leckman, J. (1976). The inheritance of affective disorders: A review of the data and hypotheses. Behavioral Genetics, 6, 227-261.

Gibbs, J. (1972). Issues in defining deviant behavior. In R.A. Scott & J.D. Douglas (Eds.), Theoretical perspectives on deviance. New York: Basic.

Goffman, E. (1961). Asylums. Garden City, NY: Doubleday Anchor.

Goldberg, L.R. (1968). Simple models or simple processes? Some research on clinical judgments. American Psychologist, 23, 483-596.

Goldberg, L.R. (1970). Man versus models of man: A rationale plus evidence for a method of improving on clinical inferences. Psychological Bulletin, 73, 422-432.

Goldberg, L.R., & Werts, C.E. (1966). The reliability of clinicians' judgments: A multitrait-multimethod approach. Journal of Consulting Psychology, 30, 199-206.

Golding, S.L., & Rorer, L.G. (1972). Illusory correlation and subjective judgment. Journal of Abnormal Psychology, 80, 249-260.

Goldman, L. (1978). Research methods for counselors: Practical approaches in field settings. New York: Wiley.

Goldstein, G., & Hersen, M. (Eds.). (1984). Handbook of psychological assessment. New York: Pergamon Press.

Goodwin, D.W., & Guze, S.B. (1984). Psychiatric diagnosis. New York: Oxford University Press.

Gove, W. (1970a). Societal reaction as an explanation of mental illness: An evaluation. American Sociological Review, 35, 873-884.

Gove. W. (1970b). Who is hospitalized: A critical evaluation of some sociological studies of mental illness. Journal of Health and Social Behavior, 11, 294-303.

Gove, W. (1975). The labelling theory of mental illness: A reply to Scheff. American Sociological Review, 40, 242-248.

Gove. W. (1979). The labeling versus the psychiatric explanation of mental illness: A debate that has become substantively irrelevant. Journal of Health and Social Behavior, 20, 301-304.

Gove, W. (Ed.). (1980). The labelling of deviance: Evaluating a perspective. London: Sage.

Gove, W. (1982). Labelling theory's explanation of mental illness: An update of recent evidence. Deviant Behavior, 3, 307-327.

Gove, W., & Fain, T. (1977). A comparison of voluntary and committed psychiatric patients. Archives of General Psychiatry, 34, 669-676.

Hamilton, D.L. (1981). Illusory correlation as a basis for stereotyping. In D.L. Hamilton (Ed.), Cognitive processes in stereotyping and intergroup behavior. Hillsdale, NJ: Erlbaum.

Hammond, K.R. (1971). Computer graphics as an aid to learning. Science, 172, 903-908.

Hansell, N. (1978). Services for schizophrenics: A lifelong approach to treatment. Hospital and Community Psychiatry, 29, 105-109.

Heider, F. (1980). The psychology of interpersonal relations. New York: Wiley.

Helson, H. (1947). Adaptation level as a frame of reference for prediction of psychophysical data. American Journal of Psychology, 60, 1-29.

Helson, H. (1964). Adaptation-level theory. New York: Harper.

Henry, K.A., & Cohen, C.I. (1983). The role of labeling processes in diagnosing borderline personality disorder. American Journal of Psychiatry, 140, 1527-1529.

Hersen, M., & Barlow, D.H. (1976). Single-case experimental designs: Strategies for studying behavior change. New York: Pergamon Press.

Heverly, M.A., Fitt, D.X., & Newman, F.L. (1984). Constructing case vignettes for evaluating clinical judgment: An empirical model. Evaluation and Program Planning, 7, 45-55.

Hines, F.R., & Williams, R.B. (1975). Dimensional diagnosis and the medical students' grasp of psychiatry. Archives of General Psychiatry, 32, 525-528.

Hirschi, T. (1973). Procedural rules and the study of deviant behavior. Social Problems, 21, 254-268.

Hoffman, P.J., Slovic, P., & Rorer, L.G. (1968). An analysis of variance model in clinical judgment. Psychological Bulletin, 69, 338-349.

Hogarth, R.M. (1975). Cognitive processes and the assessment of subjective probability distributions. Journal of the American Statistical Association, 70, 271-289.

Holsopple, J.G., & Phelan, J.G. (1954). The skills of clinicians in analysis of projective tests. Journal of Clinical Psychology, 10, 307-320.

Holt, R.R. (1961). Clinical judgment as a disciplined inquiry. Journal of Nervous and Mental Disease, 133, 369-382.

Horwitz, A. (1979). Models, muddles and mental illness labeling. Journal of Health and Social Behavior, 20, 296-300.

Huffine, C., & Clausen, J. (1979). Madness and work. Social Forces, 57, 1049-1063.

Huguenard, T., Sager, E.B., & Ferguson, L.W. (1970). Interview time, interview set, and interview outcome. Perceptual and Motor Skills, 31, 831-836.

Imershein, A., & Simons, R. (1976). Rules and examples in lay and professional psychiatry: An ethnomethodological comment on the Scheff-Gove controversy. American Sociological Review, 41, 559-563.

Johnson, D.M. (1944). Generalization of a scale of values by the averaging of practice effects. Journal of Experimental Psychology, 34, 425-436.

Johnson, T.J., Feigenbaum, R., & Weiby, M. (1964). Some determinants and consequences of the teacher's perception of causation. Journal of Educational Psychology, 55, 237-246.

Jones, E.E., & Davis, K.E. (1965). From acts to dispositions. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 2). New York: Academic Press.

Jones, E.E., Farina, A., Hastorf, A.H., Markus, H., Miller, D.T., & Scott, R.A. (1984). Social stigma: The psychology of marked relationships. New York: Basic.

Jones, E.E., & Nisbett, R.E. (1971). The actor and the observer: Divergent perceptions of the causes of behavior. New York: General Learning Press.

Jones, J.A. (1977). Self-fulfilling prophecies. Hillsdale, NJ: Erlbaum.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. Cognitive Psychology, 3, 430-454.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. Psychological Review, 80, 237-251.

Katz, D., & Braly, K.W. (1947). Verbal stereotypes and racial prejudice. In T.M. Newcombe & E.L. Hartley (Eds.), Readings in social psychology. New York: Holt.

Kaufman, E. (1979). The right to treatment suit as an agent of change. American Journal of Psychiatry, 136, 1428-1432.

Kazdin, A.E. (1980). Research design in clinical psychology. New York: Harper & Row.

Kelley, H.H. (1950). The warm-cold variable in first impressions of persons. Journal of Personality, 18, 431-439.

Kelley, H.H. (1971). Attribution in social interaction. New York: General Learning Press.

Kitsuse, J.I. (1962). Societal reaction to deviant behavior: Problems of theory and method. Social Problems, 9, 247-257.

Kitsuse, J.I., and Cicourel, A.V. (1963). A note on the uses of official statistics. Social Problems, 11, 131-139.

Klerman, G.L., & Schechter, G. (1981). The impact of psychopharmacology on the mental health service system. In W. Gove and R. Carpenter (Eds.), Nature and nurture: The fundamental connection. Lexington, KY: Lexington Books.

Kopta, S.M., Newman, F.L., McGovern, M.P., & Sandrock, D. (1986). Psychotherapeutic orientations: A comparison of conceptualizations, interventions, and treatment plan costs. Journal of Consulting and Clinical Psychology, 54, 369-374.

Kriesman, D.E., & Joy, V.D. (1974). Family response to mental illness of a relative: A review of the literature. Schizophrenia Bulletin, 10, 34-57.

Krohn, M.D., & Akers, R.L. (1977). An alternative view of the labeling versus psychiatric perspectives on societal reaction to mental illness. Social Forces, 56, 341-361.

Kruglanski, A.W., & Ajzen, I. (1983). Bias and error in human judgment. European Journal of Social Psychology, 13, 1-44.

Kulka, R.A., Veroff, J., & Donovan, E. (1979). Social class and the use of professional help for personal problems: 1957 and 1976. Journal of Health and Social Behavior, 20, 2-17.

Kurtz, R.M., & Garfield, S.L. (1978). Illusory correlation: A further exploration of Chapman's paradigm. Journal of Consulting and Clinical Psychology, 46, 1009-1015.

Kushner, K. (1978). On the external validity of two psychotherapy analogues. Journal of Consulting and Clinical Psychology, 46, 1394-1402.

Kushner, K., Bordin, E.S., & Ryan, E. (1979). Comparison of Strupp and Jenkins' audiovisual psychotherapy analogues and real psychotherapy interviews. Journal of Consulting and Clinical Psychology, 47, 765-767.

Laing, R.D. (1967). The politics of experience. New York: Ballantine.

Laing, R.D., & Esterson, A. (1964). Sanity, madness, and the family. London: Tavistock.

Langer, E.J., & Abelson, R.P. (1974). A patient by any other name: Clinician group difference in labeling bias. Journal of Consulting and Clinical Psychology, 42, 4-9.

Larsson, G., & Starrin, B. (1983). Psychiatric labels in the military setting: Some consequences on other persons' causality attribution of personal and situational factors. Scandinavian Journal of Psychology, 24, 57-61.

Lemert, E.M. (1951). Social pathology. New York: McGraw-Hill.

Lemert, E.M. (1967). Human deviance, social problems, and social control. Englewood Cliffs, NJ: Prentice-Hall.

Lerner, M.J. (1970). The desire for justice and reactions to victims. In J. Macaulay & L. Berkowitz (Eds.), Altruism and helping behavior. New York: Academic Press.

Levin, R.H. (1984). Differential anchoring effects in clinical judgment (Doctoral dissertation, State University of New York at Albany, 1984). Dissertation Abstracts International, 45, 1020B.

Levy, L. (1961). Adaptation, anchoring, and dissipation in social perception. Journal of Personality, 61, 94-104.

Lichtenstein, S., Fischoff, B., & Phillips, L.D. (1977). Calibration of probabilities: The state of the art. In H. Jungermann & G. de Zeeuw (Eds.), Decision making and change in human affairs. Dordrecht, The Netherlands: Reidel.

Light, D.W. (1982). Learning to label: The social construction of psychiatrists. In W. Gove (Ed.), Deviance and mental illness. Beverly Hills, CA: Sage.

Little, K.B., & Schneidman, E.S. (1959). Congruencies among interpretations of psychological test and anamnestic data. Psychological Monographs, 73, (6, Whole No. 476).

Lopez, S. (1984). Ethnic bias in clinical judgment: An attributional analysis (Doctoral dissertation, University of California, Los Angeles, 1984). Dissertation Abstracts International, 44, 2561B.

Lourens, P.J. (1979). All too human: Some limitations of professional judgment. Psychologia Africana, 18, 65-80.

Luborsky, L. (1962). Clinicians' judgments of mental health. Archives of General Psychiatry, 7, 407-417.

Luborsky, L., & Bachrach, H. (1974). Factors influencing clinicians' judgments of mental health: Eighteen experiences with the Health-Sickness Rating Scale. Archives of General Psychiatry, 31, 292-299.

Lueger, R.J., & Petzel, T.P. (1979). Illusory correlation in clinical judgment: Effects of amount of information to be processed. Journal of Consulting and Clinical Psychology, 47, 1120-1121.

Lyon, D., & Slovic, P. (1976). Dominance of accuracy information and neglect of base rates in probability estimation. Acta Psychologica, 40, 287-298.

Manning, P.K. (1973). On deviance. Contemporary Sociology, 2, 123-128.

May, P.R., & Tuma, A.H. (1964). Choice of criteria for the assessment of treatment outcome. Journal of Psychiatric Research, 2, 199-209.

Meehl, P.E. (1954). Clinical versus statistical prediction. Minneapolis: University of Minnesota Press.

Meehl, P.E. (1956). Wanted--a good cookbook. American Psychologist, 11, 263-272.

Meehl, P.E. (1957). When shall we use our heads instead of the formula? Journal of Counseling Psychology, 4, 268-273.

Meehl, P.E. (1959). Some ruminations of the validation of clinical procedures. Canadian Journal of Psychology, 13, 102-128.

Meehl, P.E. (1960). The cognitive activity of the clinician. American Psychologist, 15, 19-27.

Mensh, I.N., & Wishner, J. (1947). Asch on "Forming impressions of personality": Further evidence. Journal of Personality, 16, 188-191.

Mischel, W. (1968). Personality and assessment. New York: Wiley.

Morrissey, J., Tessler, R., & Farrin, L. (1979). Being seen but not admitted: A note on some neglected aspects of state hospital deinstitutionalization. American Journal of Orthopsychiatry, 49, 153-156.

Myers, J., Lindenthal, J., & Pepper, M. (1975). Life events, social integration and psychiatric symptomatology. Journal of Health and Social Behavior, 16, 421-427.

Nathan, P.E., Zare, N.C., Simpson, H.F., & Ardberg, M.M. (1969). A systems analytic model of diagnosis: The diagnostic validity of abnormal psychomotor behavior. Journal of Clinical Psychology, 25, 3-9.

Neisser, U. (1967). Cognitive psychology. New York: Appleton-Century-Crofts.

Nieva, V.F., & Gutek, B.A. (1980). Sex effects on evaluation. Academy of Management Review, 5, 267-276.

Nisbett, R.E., & Borgida, E. (1975). Attribution and the psychology of prediction. Journal of Personality and Social Psychology, 32, 932-943.

Nisbett, R.E., Borgida, E., Crandall, R., & Reed, H. (1976). Popular induction: Information is not necessarily informative. In J.S. Carroll & J.W. Payne (Eds.), Cognition and social behavior. Hillsdale, NJ: Erlbaum.

Nisbett, R.E., & Ross, L. (1980). Human inference: Strategies and shortcomings of social judgment. Englewood Cliffs, NJ: Prentice-Hall.

Nunnally, J.C. (1961). Popular conceptions of mental health. New York: Holt, Rinehart, & Winston.

Olmstead, D., & Durham, K. (1976). Stability of mental health attitudes: A semantic differential study. Journal of Health and Social Behavior, 17, 35-44.

Olshansky, S., Brob, S., & Dahl, M.K. (1960). Survey of employment experiences of patients discharged from three state mental hospitals during the period 1951-1953. Mental Hygiene, 44, 510-521.

Oskamp, S. (1965). Overconfidence in case-study judgments. Journal of Consulting Psychology, 29, 261-265.

Oskamp, S. (1967). Clinical judgment from the MMPI: Simple or complex. Journal of Clinical Psychology, 23, 411-415.

Patterson, D.R. (1983). Socio-cognitive biases in clinical judgment (Doctoral dissertation, Florida State University, 1982). Dissertation Abstracts International, 43, 2290A.

Peak, H. (1958). Psychological structure and psychological activity. Psychological Review, 65, 325-347.

Pepinsky, H.B., & Pepinsky, P. (1954). Counseling theory and practice. New York: Ronald Press.

Perloff, R.M., Padgett, V.R., & Brock, T.C. (1980). Socio-cognitive biases in the evaluation process. In R. Perloff & B. Perloff (Eds.), Values, ethics, and standards in evaluation. San Francisco: Jossey-Bass.

Phelan, J.G. (1964). Rationale employed by clinical psychologists in diagnostic judgment. Journal of Clinical Psychology, 20, 454-458.

Rabkin, J. (1974). Public attitudes toward mental illness: A review of the literature. Schizophrenia Bulletin, 19, 9-33.

Rabkin, J. (1979). Who is called mentally ill: Public and professional views. Journal of Community Psychiatry, 7, 253-258.

Reed, P.L., & Jackson, D.N. (1975). Clinical judgment of psychopathology: A model for inferential accuracy. Journal of Abnormal Psychology, 84, 475-482.

Regier, D., Goldberg, I.D., & Taube, C.A. (1978). The de facto U.S. mental health service: A public health perspective. Archives of General Psychiatry, 35, 685-693.

Richards, J.M. (1963). Reconceptualization of the clinical and statistical prediction controversy in terms of components of accuracy of interpersonal perception scores. Psychological Reports, 12, 443-448.

Rosenhan, D.L. (1973). On being sane in insane places. Science, 179, 250-257.

Ross, L. (1977). The intuitive psychologist and his shortcomings. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 10). New York: Academic Press.

Ross, L., Lepper, M.R., & Hubbard, M. (1975). Perseverance in self perception and social perception: Biased attributional processes in the debriefing paradigm. Journal of Personality and Social Psychology, 32, 880-892.

Ross, L., Lepper, M.R., Strack, F., & Steinmetz, J. (1977). Social explanation and social expectation: Effects of real and hypothetical explanations on subjective likelihood. Journal of Personality and Social Psychology, 35, 817-829.

Rothaus, P., Hanson, P.G., Cleveland, S.E., & Johnson, D.L. (1963). Describing hospitalization: A dilemma. American Psychologist, 18, 85-89.

Rubin, M., & Shontz, F.C. (1960). Diagnostic prototypes and diagnostic processes of clinical psychologists. Journal of Consulting Psychology, 24, 234-239.

Sarbin, T. (1967). The scientific status of the mental illness concept. In S. Plog (Ed.), Determinants of mental illness: A handbook. New York: Holt, Rinehart, & Winston.

Sarbin, T. (1972). Stimulus/response: Schizophrenia is a myth, born of metaphor, meaningless. Psychology Today, 6(1), 18-27.

Sawyer, J. (1966). Measurement and prediction, clinical and statistical. Psychological Bulletin, 66, 178-200.

Scheff, T. (1964). The societal reaction to deviance: Ascriptive elements in the psychiatric screening of mental patients in a midwestern state. Social Problems, 11, 401-413.

Scheff, T. (1966). Being mentally ill: A sociological theory. Chicago: Aldine.

Scheff, T. (1974). The labeling theory of mental illness. American Sociological Review, 39, 444-452.

Scheff, T. (1975). Reply to Chauncey and Gove. American Sociological Review, 40, 252-257.

Scheff, T. (1976). Reply to Imershein and Simons. American Sociological Review, 41, 563-564.

Scheff, T. (1979). Reply to comment by Horwitz. Journal of Health and Social Behavior, 20, 305-306.

Schlesinger, G.H. (1984). The effects of a self-analytic exercise on clinical judgment (Doctoral dissertation, New York University, 1983). Dissertation Abstracts International, 44, 2259B.

Scull, A. (1984). Competing perspectives on deviance. Deviant Behavior, 5, 275-289.

Sherif, M., & Hovland, C.I. (1961). Social judgment. New Haven: Yale University Press.

Siassi, I. (1984). Psychiatric interviews and mental status examinations. In G. Goldstein & M. Hersen (Eds.), Handbook of psychological assessment. New York: Pergamon Press.

Sines, L.K. (1959). The relative contribution of four kinds of data to accuracy in personality assessment. Journal of Consulting Psychology, 23, 483-492.

Skodol, A., & Spitzer, R.L. (1982). DSM-III: Rationale, basic concepts, and some differences from ICD-9. Acta Psychiatrica Scandinavia, 66, 271-281.

Slovic, P., & Fischoff, B. (1977). On the psychology of experimental surprises. Journal of Experimental Psychology: Human Perception and Performance, 3, 544-551.

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 6, 649-744.

Smedslund, J. (1966). Note on learning, contingency, and clinical experience. Scandinavian Journal of Psychology, 7, 265-266.

Spitzer, R.L. (1976). More on pseudoscience in science and the case for psychiatric diagnosis. Archives of General Psychiatry, 33, 459-470.

Spitzer, R.L., Forman, J.B., & Nee, J. (1979). DSM-III field trials: Initial interrater diagnostic reliability. American Journal of Psychiatry, 136, 815-817.

Spitzer, R.L., Gibbon, M., & Endicott, J. (1971). Family Evaluation Form. New York: Biometrics Research.

Spitzer, R.L., Skodol, A., Gibbon, M. & Williams, J. (1981). DSM-III case book. Washington, DC: American Psychiatric Association.

Spitzer, R.L., Skodol, A., Gibbon, M., & Williams, J. (1983). Psychopathology: A case book. New York: McGraw-Hill.

Strasburger, E.L., & Jackson, D.N. (1977). Improving accuracy in a clinical judgment task. Journal of Consulting and Clinical Psychology, 45, 303-309.

Strohmer, D.C., Haase, R.F., Biggs, D.A., & Keller, K.E. (1982). Process models of clinical judgment. Journal of Counseling Psychology, 29, 597-606.

Summers, D.A., Taliaferro, D.J., & Fletcher, D.J. (1969). Subjective vs. objective description of judgment policy. Psychonomic Science, 18, 249-250.

Szasz, T. (1961). The myth of mental illness. New York: Hoeber-Harper.

Szasz, T. (1970). The manufacture of madness. New York: Harper & Row.

Taft, R. (1955). The ability to judge people. Psychological Bulletin, 52, 1-23.

Tagiuri, R., & Petrullo, L. (1958). Person perception and interpersonal behavior. Palo Alto, CA: Stanford University Press.

Tannenbaum, F. (1938). Crime and the community. Boston: Ginn.

Taylor, C.B. (1983). DSM-III and behavioral assessment. Behavioral Assessment, 5, 5-14.

Temerlin, M.K. (1968). Suggestion effects in psychiatric diagnosis. Journal of Nervous and Mental Disease, 147, 349-353.

Temerlin, M.K. (1970). Diagnostic bias in community mental health. Community Mental Health Journal, 6, 110-117.

Trute, B., & Loewen, A. (1978). Public attitude toward the mentally ill as a function of prior personal experience. Social Psychiatry, 13, 79-84.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. Psychological Bulletin, 76, 105-110.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 5, 207-232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-1131.

Volkmann, J. (1951). Scales of judgment and their implications for social psychology. In J.H. Rorer & M. Sherif (Eds.), Social psychology at the crossroads. New York: Harper.

Wallach, M.S., & Schoof, K. (1965). Reliability of degree of disturbance rating. Journal of Clinical Psychology, 21, 271-275.

Walster, E. (1966). Assignment of responsibility for an accident. Journal of Personality and Social Psychology, 3, 73-79.

Wechsler, H., Solomon, L., & Kramer, B.N. (Eds.). (1970). Social psychology and mental health. New York: Holt, Rinehart, & Winston.

Weinstein, R. (1982). The mental hospital from the patient's view: Qualitative vs. quantitative data. In W. Gove (Ed.), Deviance and mental illness. New York: Sage.

Wiggins, J.S. (1973). Personality and prediction: Principles of personality assessment. Cambridge, MA: Addison-Wesley.

Wiggins, J.S. (1981). Clinical and statistical prediction: Where are we and where do we go from here? Clinical Psychology Review, 1, 3-18.

Wishner, J. (1960). Reanalysis of "Impressions of personality". Psychological Review, 67, 96-112.

Wortis, J. (1982). DSM-III: The big debate. Biological Psychiatry, 17, 1363-1365.

BIOGRAPHICAL SKETCH

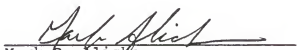
Curtis Craig Walling was born August 14, 1959, to Craig and Kathy Walling of Phoenix, Arizona. He graduated from Phoenix Alhambra High School with high honors in 1977. Athletic scholarships led him in 1977 to Scottsdale Community College and in 1978 to Glendale Community College, where he was also admitted to the Phi Theta Kappa scholastic fraternity. In 1979 he accepted a scholarship to attend Arizona State University, where he studied psychology. He transferred the following year to Brigham Young University, where he earned the degree of Bachelor of Science in psychology in 1982.

In August, 1982, Curtis enrolled at the University of Florida. He earned the degree of Master of Science in psychology in April, 1984, and has since that time worked towards completion of the requirements for the degree of Doctor of Philosophy in counseling psychology.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Greg J. Neimeyer, Chairman
Associate Professor of
Psychology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



Mark D. Alicke
Assistant Professor of
Psychology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

James Archer, Jr.
Professor of Psychology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Paul W. Fitzgerald
Professor of Counselor Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Mary A. Fukuyama
Assistant Professor of Psychology

This dissertation was submitted to the Graduate Faculty of the Department of Psychology in the College of Liberal Arts and Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree Doctor of Philosophy.

May 1987

Dean, Graduate School